



***Automatisk litelogianalyse basert på
maskinlæringsalgoritmer med
utgangspunkt i Naive Bayes Classifiser***

av

**Roar Hauger
Nils-Edvard Lileng Holene**

**Masteroppgave i
informasjons- og kommunikasjonsteknologi**

**Høgskolen i Agder
Fakultet for teknologi**

**Grimstad
mai 2005**

Sammendrag

Når det skal bores etter olje eller andre mineraler vil det være en fordel å vite noe om grunnen det skal bores i. Ved hjelp av sensorer som sitter på borestrengen kan man med stor sikkerhet avgjøre i hvilken litelogi det blir boret. Med en litelogi menes det et fast stoff som for eksempel Skifer, Sand eller Stein. Når man vet hvilke litelogier grunnen består av, vil sannsynligheten for å finne olje og mineraler bli avslørt. De målte verdiene $G_{c,d}$ fra sensorene blir lagret i en logg, og loggen blir sendt videre til en geolog som utfører tolkningen av denne.

Ved å utvikle en automatisk loggtolker vil tolkningen av loggene effektiviseres ved at loggene ikke trengs å sendes til en geolog. En datamaskin vil klassifisere de forskjellige litelogiene ved hjelp av verdiene i loggen, og litelogien blir analysert fortløpende. Dette vil føre til hurtigere avgjørelser under boringen som resulterer i tid og kostnadsbesparelse, da det koster mange tusen kroner per dag å holde en boreplattform operativ.

Et problem med å utvikle en automatisk loggtolker er at sensorene er plaget av støy, noe som gjør at verdier $G_{c,d}$ målt fra samme litelogi varierer. Videre er det et problem at antall sensorer ikke er fast eller alltid de samme i loggene. Under boringen blir det målt store mengder data fra sensorene.

På bakgrunn av de overnevnte problemene ble det undersøkt hvilke metoder som ville egne seg best til å utvikle en loggtolker. Vi valgte å benytte en metode som bygger på normalfordeling da en normalfordeling tillater variasjoner i verdiene som representerer en litelogi. Metoden vi valgte å undersøke nærmere var Naive Bayes Classifier. Denne metoden har fordelen at den tolererer støy i datasettet samtidig som den gir en gradering av klassifiseringen. Metoden kan også trene opp systemet inkrementelt. Det vil si at det ikke kreves at alle treningsdataene blir lest inn på en gang, og dataene kan derfor trenes opp over en lengre tidsperiode. Den benytter flere sensorer til å klassifisere litelogiene som er med på å redusere feil.

Fire forskjellige logger ble brukt for å teste ut metoden, og loggene ble valgt da de gav et representativt utvalg av hvordan grunnen kan se ut. Det ble gjennomført en sammenlikning mellom Naive Bayes Classifier og en klassifisering der én sensor ble benyttet. Det viste seg at Naive Bayes Classifier klassifiserte loggene bedre i tre av fire tilfeller.

Det er naturlig å tro at det eksisterer en sammenheng mellom etterfølgende målinger \vec{G}_d . Om det kan dras nytte av denne korrelasjonen for å gi økt klassifiseringsnøyaktighet ble undersøkt ved hjelp av Hidden Markov Model. Den har de samme egenskapene som Naive Bayes Classifier, men i tillegg tar den hensyn til før og etterliggende målinger \vec{G}_d i tid / rom. Sammenlikningen mellom Naive Bayes Classifier og Hidden Markov Model, viste at det fantes en korrelasjon mellom målingene \vec{G}_d . Hidden Markov Model benyttet seg av denne korrelasjonen til å klassifisere bedre enn Naive Bayes Classifier i tre av de fire loggene.

En undersøkelse om det er mulig å oppnå en bedre klassifisering ved å optimalisere Naive Bayes Classifier og Hidden Markov Model er blitt utført. Ved å fjerne 1 % av de såkalte uteliggerne viste det seg at klassifiseringen ble like bra eller bedre i alle fire loggene. Vi fant ut at alle tilgjengelige sensorer bør benyttes i klassifiseringsmetodene.

Etter flere undersøkelser viste det seg at verdiområdet, som kjennetegner hver litelogi, forandrer seg avhengig av hvor det blir boret.

Ut fra resultatene i denne rapporten ser det ut til at Hidden Markov Model bør bli benyttet som klassifiseringsmetode. Alle tilgjengelige sensorer bør benyttes, og 1 % av uteliggerne bør fjernes. Systemet må trenes opp med data fra samme borehull for å oppnå høyest grad av riktig klassifisering. Om resultatene ønskes i "sanntid" mens boringen pågår, vil Naive Bayes Classifier være et godt alternativ til Hidden Markov Model. Eventuelt kan Hidden Markov Model konfigurere slik at den tilnærmet klassifiserer i "sanntid" ved for eksempel å benytte kun fem etterliggende målinger \vec{G}_d .

Forord

Denne rapporten er skrevet som en avsluttende del av den 2-årige masterutdanningen innen informasjons- og kommunikasjonsteknologi (IKT) ved Høgskolen i Agder, avdeling for teknologi i Grimstad. Arbeidet har pågått over 20 uker fra januar til mai 2005 og har blitt utført ved Høgskolen i Agder, i dialog med Sense Intellifield i Kristiansand. Prosjektet tilsvarer 30 studiepoeng.

Oppgaven Automatisk litelogianalyse basert på maskinlæringsalgoritmer med utgangspunkt i Naive Bayes Classifier er gitt av Sense Intellifield AS.

Vi vil gjerne få takke våre veiledere Ole-Christoffer Granmo og eksterne veileder Arne Wiklund fra Sense Intellifield for veiledning og assistanse gjennom prosjektperioden.

Grimstad, mai 2005

Roar Hauger og Nils-Edvard Lileng Holene

Innholdsfortegnelse

| | |
|---|------------|
| SAMMENDRAG | I |
| FORORD | III |
| INNHOLDSFORTEGNELSE | IV |
| FIGURLISTE | VI |
| TABELLISTE | VII |
| 1 INNLEDNING | 1 |
| 1.1 BAKGRUNN | 1 |
| 1.2 OPPGAVEBESKRIVELSE | 1 |
| 1.3 BEGRENSNINGER | 2 |
| 1.4 RELATERT ARBEID | 2 |
| 1.5 RAPPORTENS OPPBYGNING | 3 |
| 2 BESKRIVELSE AV DATASETET | 5 |
| 2.1 BESKRIVELSE AV LOGGENE SOM ER BRUKT | 5 |
| 2.2 FORKLARING AV SENSORENE | 8 |
| 2.3 FORKLARING AV LITELOGIENE..... | 10 |
| 2.4 POTENSIELLE PROBLEMER MED LOGGENE..... | 11 |
| 3 BAYESIANSK BESLUTNINGSTEORI | 13 |
| 3.1 BAYES TEOREM | 13 |
| 3.2 NORMALFORDELING | 14 |
| 3.3 MIDDELVERDI..... | 15 |
| 3.4 STANDARDAVVIK | 15 |
| 3.5 KLASSIFISERING VED HJELP AV ÉN SENSOR..... | 16 |
| 3.6 NAIVE BAYES CLASSIFIER | 17 |
| 3.6.1 <i>Klassifisering av flere sensorer</i> | 17 |
| 3.7 HIDDEN MARKOV MODEL | 18 |
| 3.8 FORWARD – BACKWARD PROSEDYREN | 21 |
| 3.8.1 <i>Forward algoritmen</i> | 21 |
| 3.8.2 <i>Backward algoritmen</i> | 22 |
| 4 EKSPERIMENTER | 24 |
| 4.1 LEAVE-ONE-OUT | 24 |
| 4.2 KLASSIFISERING VED HJELP AV ÉN SENSOR | 25 |
| 4.2.1 <i>Logg 1</i> | 25 |
| 4.2.2 <i>Logg 2</i> | 26 |
| 4.2.3 <i>Logg 3</i> | 27 |
| 4.2.4 <i>Logg 4</i> | 28 |
| 4.2.5 <i>Verdiområdet ved boreddybde</i> | 28 |
| 4.3 KLASSIFISERING VED HJELP AV NAIVE BAYES CLASSIFIER..... | 29 |
| 4.3.1 <i>Logg 1</i> | 30 |
| 4.3.2 <i>Logg 2</i> | 30 |
| 4.3.3 <i>Logg 3</i> | 31 |
| 4.3.4 <i>Logg 4</i> | 32 |
| 4.4 KLASSIFISERING VED HJELP AV HIDDEN MARKOV MODEL..... | 32 |
| 4.4.1 <i>Logg 1</i> | 32 |
| 4.4.2 <i>Logg 2</i> | 33 |
| 4.4.3 <i>Logg 3</i> | 33 |
| 4.4.4 <i>Logg 4</i> | 34 |
| 4.5 UTBEDRINGER AV KLASSIFISERINGEN..... | 35 |
| 4.5.1 <i>Fjerne uteliggere</i> | 35 |
| 4.5.2 <i>Fjerne sensorer</i> | 39 |
| 4.6 TRENE PÅ TVERS AV LOGGENE..... | 42 |

| | | |
|----------|--|-----------|
| 5 | DISKUSJON AV RESULTATER..... | 46 |
| 5.1 | ÉN SENSOR..... | 46 |
| 5.2 | NAIVE BAYES CLASSIFIER..... | 47 |
| 5.3 | HIDDEN MARKOV MODEL..... | 47 |
| 5.4 | UTBEDRINGER AV KLASSIFIKATORENE..... | 48 |
| 5.5 | TRENE PÅ TVERS AV LOGGER..... | 49 |
| 5.6 | VIDERE ARBEID..... | 49 |
| 6 | KONKLUSJON..... | 50 |
| | REFERANSER..... | 52 |
| | VEDLEGG..... | 53 |
| | VEDLEGG A – FORVIRRINGSMATRISER FOR LOGG 1 | |
| | VEDLEGG B – FORVIRRINGSMATRISER FOR LOGG 2 | |
| | VEDLEGG C – FORVIRRINGSMATRISER FOR LOGG 3 | |
| | VEDLEGG D – FORVIRRINGSMATRISER FOR LOGG 4 | |

Figurliste

| | |
|---|----|
| Figur 2.1: Eksempel på hvordan litelogiene blir symbolisert i en logg | 10 |
| Figur 2.2: Eksempel på hvordan en logg ser ut..... | 11 |
| Figur 3.1: Eksempel på en normalfordeling..... | 14 |
| Figur 3.2: Eksempel på en overgangsmatrise..... | 19 |
| Figur 3.3: Eksempel på initialmatrise og overgangssannsynlighet..... | 20 |
| Figur 3.4: Eksempel på Forward – Backward..... | 21 |
| Figur 4.1: Eksempel på forvirringsmatrise..... | 25 |
| Figur 4.2: Verdiområdet ved dybde | 29 |
| Figur 4.3: Eksempel på målinger over flere dimensjoner | 29 |
| Figur 4.4: Utsnitt av en logg som viser innslag av sedimenter | 35 |
| Figur 4.5: Antall sensorer man bør benytte på Logg 1..... | 40 |
| Figur 4.6: Antall sensorer man bør benytte på Logg 2..... | 41 |
| Figur 4.7: Illustrering av fordelingen til Neutron Porosity over alle loggene..... | 42 |
| Figur 4.8: Illustrering av den nye fordelingen til Neutron Porosity over alle loggene. | 43 |

Tabelliste

| | |
|--|----|
| Tabell 2.1: Sensorer i Logg 1 | 6 |
| Tabell 2.2: Sensorer i Logg 2 | 6 |
| Tabell 2.3: Sensorer i Logg 3 | 7 |
| Tabell 2.4: Sensorer i Logg 4 | 7 |
| Tabell 4.1: Forvirringsmatrise for den beste sensoren i Logg 1..... | 26 |
| Tabell 4.2: Forvirringsmatrise for den dårligste sensoren i Logg 1 | 26 |
| Tabell 4.3: Forvirringsmatrise for den beste sensoren i Logg 2..... | 26 |
| Tabell 4.4: Forvirringsmatrise for den dårligste sensoren i Logg 2 | 27 |
| Tabell 4.5: Forvirringsmatrise for den beste sensoren i Logg 3..... | 27 |
| Tabell 4.6: Forvirringsmatrise for den dårligste sensoren i Logg 3 | 27 |
| Tabell 4.7: Forvirringsmatrise for den beste sensoren i Logg 4..... | 28 |
| Tabell 4.8: Forvirringsmatrise for den dårligste sensoren i Logg 4..... | 28 |
| Tabell 4.9: Forvirringsmatrise for Naive Bayes Classifier i Logg 1 | 30 |
| Tabell 4.10: Forvirringsmatrise for Naive Bayes Classifier i Logg 2..... | 30 |
| Tabell 4.11: Forvirringsmatrise for Naive Bayes Classifier i Logg 3..... | 31 |
| Tabell 4.12: Forvirringsmatrise for Naive Bayes Classifier i Logg 4..... | 32 |
| Tabell 4.13: Forvirringsmatrise for Hidden Markov Model på Logg 1 | 32 |
| Tabell 4.14: Forvirringsmatrise for Hidden Markov Model på Logg 2 | 33 |
| Tabell 4.15: Forvirringsmatrise for Hidden Markov Model på Logg 3 | 33 |
| Tabell 4.16: Forvirringsmatrise for Hidden Markov Model på Logg 4 | 34 |
| Tabell 4.17: Logg 1 klassifisert med Naive Bayes Classifier uten 1 % av uteliggerne | 36 |
| Tabell 4.18: Logg 1 klassifisert med Hidden Markov Model uten 1 % av uteliggerne | 36 |
| Tabell 4.19: Logg 2 klassifisert med Naive Bayes Classifier uten 1 % av uteliggerne | 37 |
| Tabell 4.20: Logg 2 klassifisert med Hidden Markov Model uten 1 % av uteliggerne | 37 |
| Tabell 4.21: Logg 1 klassifisert med Naive Bayes Classifier uten 1 ‰ av uteliggerne..... | 37 |
| Tabell 4.22: Logg 1 klassifisert med Hidden Markov Model uten 1 ‰ av uteliggerne | 38 |
| Tabell 4.23: Logg 2 klassifisert med Naive Bayes Classifier uten 1 ‰ av uteliggerne..... | 38 |
| Tabell 4.24: Logg 2 klassifisert med Hidden Markov Model uten 1 ‰ av uteliggerne | 38 |
| Tabell 4.25: Liste over hvilke sensorer som skiller litelogiene best | 39 |
| Tabell 4.26: Logg 1 klassifisert med Naive Bayes Classifier med nytt treningsgrunnlag | 44 |
| Tabell 4.27: Logg 1 klassifisert med Hidden Markov Model med nytt treningsgrunnlag | 44 |
| Tabell 4.28: Logg 1 klassifisert med Hidden Markov Model, nytt treningsgrunnlag og snittovergangsmatrise..... | 45 |

1 Innledning

I løpet av det siste tiåret har det blitt mer vanlig å benytte datamaskiner for å ta bedre og mer nøyaktige beslutninger. En beslutning som kunne tatt lang tid for et menneske, vil kunne bli utført mye raskere av en datamaskin. Etter hvert som datamaskinen har fått økt kapasitet, har det også blitt mulig å gjøre flere krevende oppgaver og benytte seg av mer avanserte metoder. Denne masteroppgaven vil bruke datamaskiner til å bestemme hva slags litologi det bores i på bakgrunn av informasjon om de forskjellige litologiene. Med en litologi menes det et fast stoff som for eksempel Skifer, Sand eller Stein.

1.1 Bakgrunn

Ved boring etter olje eller andre mineraler er det viktig å utføre analyser av grunnen det skal bores i. Analysene vil fortelle hvor det er mest sannsynlig å finne oljen eller mineralene det bores etter.

Vi innfører noen matematiske begreper for å lettere kunne skille mellom begrepene sensor, en verdi målt av en sensor og verdier målt av flere sensorer ved en dybde. Sensorene betegnes som et sett $C = \{c_1, c_2, \dots, c_n\}$ hvor c_i er en gitt sensor, $c_i \in C$. En verdi målt av en sensor ved én dybde d betegnes $G_{c,d}$, og denne blir kalt en verdi i resten av rapporten. Alle verdiene målt av sensorene ved en dybde d betegnes som \vec{G}_d hvor $\vec{G}_d = [G_{c_1,d}, G_{c_2,d}, \dots, G_{c_n,d}]$, og denne blir kalt en måling i resten av rapporten.

Under boringen logges det verdier $G_{c,d}$ fra flere sensorer som forteller om egenskapene til litologiene. Loggene blir deretter sendt til en geolog som utfører en analyse av de lagrede verdiene $G_{c,d}$. Geologen sender tilbake en litologiprofil [12] som forteller hvilke litologier det har blitt boret i. Analysen som geologen utfører, blir gjort for hånd. Det kan være flere tusen målinger \vec{G}_d i hver logg, og analysen vil naturligvis ta tid.

Det vil være hensiktsmessig å få utført analysen av loggen på boreplattformen mens boringen pågår. På denne måten slipper man å vente på å få analysert loggen manuelt, noe som kan føre til sparing av både tid og penger. Å få litologiene analysert fortløpende under boring, vil føre til at boringen blir utført på en mer effektiv måte. Om for eksempel boringen begynner å gå tregt vil det være mulig å se om årsaken til dette er at litologien det bores i er hard, eller om borekronen er sløv og må skiftes.

En mulig løsning på problemet vil være å utføre en analyse av litologiene ved bruk av datamaskiner og maskinlæring. Det er dette som er hovedmålsetningen for masteroppgaven.

1.2 Oppgavebeskrivelse

Basert på semitolkede logger fra boreplattformer skal vi konstruere et system som skal kunne foreslå hva slags litologi det bores i. De semitolkede loggene inneholder verdier $G_{c,d}$ målt av forskjellige sensorer i borestrengen annotert med litologien foreslått av en geolog. Med andre ord skal systemet trenes opp til å foreslå det samme som en geolog ville gjort, basert på målte verdier $G_{c,d}$ i loggene.

Det skal undersøkes om Naive Bayes Classifier og Hidden Markov Model egner seg til en slik litelogianalyse. Fordelen ved bruk av Naive Bayes Classifier i problemstillingen skissert over, er at den tolererer støy i datasettet. Samtidig gir den en gradering av hvor sikker klassifiseringen er. En annen fordel med Naive Bayes Classifier er at systemet trenes opp inkrementelt. Det vil si at metoden ikke krever at alle treningsdataene blir lest inn på en gang, men kan trenes opp over en lengre tidsperiode. Videre er minne og prosesseringsforbruk konstant. Den andre modellen som skal undersøkes er Hidden Markov Model. Denne bygger på Naive Bayes Classifier, og har derfor de samme fordelene. Forskjellen er at den i tillegg tar hensyn til før og etterfølgende målinger \vec{G}_a i tid / rom.

Naive Bayes Classifier og Hidden Markov Model skal sammenlignes, hovedsaklig for å undersøke i hvilken grad det er korrelasjon mellom etterfølgende målinger \vec{G}_a . Det skal også undersøkes i hvilken grad denne korrelasjonen kan dras nytte av i Hidden Markov Model for å gi økt klassifiseringsnøyaktighet.

Til slutt skal det ses på mulige forbedringer eller utvidelser av Naive Bayes Classifier og Hidden Markov Model med tanke på særtrekkene ved litelogianalyse. Et spesielt hensyn som må tas i betraktning er om verdiområdene, som kjennetegner hver litelogi, forandrer seg avhengig av borested og boreddybde. Et annet hensyn som må gjøres er at loggene ikke nødvendig er like fra borested til borested.

1.3 Begrensninger

Fokus i dette prosjektet har ikke vært å lage et system som skal kunne tas i bruk med umiddelbar virkning, men å undersøke hvor godt metodene Naive Bayes Classifier og Hidden Markov Model egner seg til å klassifisere litelogiene. Undersøkelsene kan derimot benyttes til å utvikle en loggertolker som bygger på disse metodene.

En annen begrensning som oppstår ved å bruke Hidden Markov Model er at ikke svaret vil være i "sanntid". Årsaken til dette er at en av grunntankene bak metoden er at det er korrelasjon mellom målingene \vec{G}_a . Metoden er dermed avhengig av målingene \vec{G}_a gjort før og etter, slik at man kan klassifisere til riktig litelogi. Dersom det viser seg at Hidden Markov Model klassifiserer litelogiene best, kan metoden konfigureres slik at klassifiseringen skjer tilnærmet i "sanntid". Dette kan for eksempel gjøres ved å kun benytte fem etterfølgende målinger \vec{G}_a i stedet for alle etterfølgende målinger \vec{G}_a .

1.4 Relatert arbeid

Under relatert arbeid blir noen andre rapporter og arbeid innenfor samme fagområde presentert. Dette vil gjøre det lettere å sette seg inn i arbeidsområdet og forstå emnet som helhet.

Før man ser på relatert arbeid vil det være interessant å se på bakgrunnsinformasjon. Et eksempel på dette er i hvilket verdiområde litelogiene ligger, og på hvilket format verdiene $G_{c,d}$ lagres i loggen. Mer informasjon finnes under [11].

Et annet sentralt område innenfor dette prosjektet er å se på hvordan de tolkede loggene blir illustrert. I [12] vises det hvordan en litelogiprofil kan bli laget og illustrert, se Kapittel 2.3 for et eksempel på et utsnitt av en slik litelogiprofil.

Det er blitt utarbeidet flere rapporter angående litelogianalyse hvor det er forsøkt å klassifisere litelogiene ved hjelp av ulike sensorer og metoder.

En rapport som kan være interessant for leseren som komplement til denne rapporten er [7]. Denne rapporten undersøker hvor godt man kan klassifisere litelogiene ved å benytte seg av sensorene Thermal Neutron Porosity og High Resolution Gamma Ray. Programmet LogTrans [9], som er utviklet av Cooperative Research Centre for Mining Technology & Equipment, blir benyttet til å skille litelogiene. Dette er et softwareprogram for automatisk tolkning av borehull for geofysiske logger. I LogTrans trener man opp systemet ved å finne middelveid og standardavvik for deretter å benytte nærmeste-nabo-metoden til å klassifisere litelogiene.

Rapport [8] viser en annen metode som kan benyttes til å utføre en litelogianalyse. I denne rapporten blir litelogiene klassifisert ved hjelp av Backpropagation Neural Network og Fuzzy Associative Memory. Problemet med bruk av nevralt nettverk er at de benytter en såkalt sort boks metode. Med dette menes det at man har verdier $G_{c,d}$ som input og får et resultat som output. Det er ingen mulighet til å undersøke hvilke beregningene som gjøres for å komme fram til dette resultatet. Et annet problem med Backpropagation Neural Networks er at den ikke har mulighet til å benytte seg av før og etterliggende målinger \vec{G}_d . En annen årsak til at vi valgte å undersøke en annen metode var at en av grunnideene i denne metoden er å benytte geologenes kunnskap om litelogiene. Vi hadde ikke tilgang til en geolog som kunne sørge for denne kunnskapen.

1.5 Rapportens oppbygning

Vi gjorde en analyse av loggene som ble logget under boring av oljebrønner. I Kapittel 2 kommer en innføring i loggene som har blitt lagt til grunn for dette prosjektet. Her blir de og innholdet beskrevet. Det blir også satt fokus på problemer som vil kunne oppstå med disse loggene.

Hovedfokus i denne oppgaven har vært å undersøke hvor godt Naive Bayes Classifier og Hidden Markov Model egner seg i et slikt klassifiseringsproblem. I Kapittel 3 er disse metodene samt andre metoder som benyttes i dette prosjektet beskrevet.

I kapittel 4 blir eksperimenter og resultatene av disse beskrevet. Vi utførte en analyse av de forskjellige sensorene som ble registrert i de forskjellige loggene. Grunnen til dette var å se hvilke sensorer som egner seg til å skille litelogiene, og hvilke som eventuelt ikke egner seg.

Det ble også utført et eksperiment hvor vi så på hvor godt Naive Bayes Classifier klassifiserte litelogiene, ved hjelp av flere sensorer. Eksperimentet ble gjort for å undersøke om dette ville føre til en bedre klassifisering. Deretter ble Hidden Markov Model brukt for å undersøke om det fantes en korrelasjon mellom målingene \vec{G}_d . I tillegg har vi også undersøkt hvor stor rolle overgangsmatrisen spiller når litelogiene klassifiseres ved hjelp av Hidden Markov Model. Se Kapittel 3.7 for nærmere forklaring av hvordan overgangsmatrisen blir benyttet i Hidden Markov Model.

Det er i tillegg blitt utført undersøkelser om klassifiseringen kan forbedres ved å fjerne eventuelle uteliggere, dette gjorde vi for å fjerne feil som kan oppstå grunnet sedimenter. Kapittel 4.5.1 vil forklare problemene rundt sedimenter nærmere. Vi har også undersøkt om klassifiseringen vil kunne forbedres ved å utelukke en eller flere sensorer.

Vi forsøkte å klassifisere en logg ved å trene opp klassifiseringsmetodene med målinger \vec{c}_d fra andre logger. Dette gjorde vi for å undersøke om verdiområdene, som kjennetegner hver litelogi, forandrer seg avhengig av hvor det bores. Dersom verdiområdet er konstant for hver litelogi, vil systemet kunne trenes opp uavhengig av hvor boringen finner sted. Om så ikke er tilfelle må man ta hensyn til hvor det bores og innhente informasjon og oppdatere treningsdataene. Det er også mulig at klassifiseringsmetodene må trenes opp på nytt med data fra brønnen der det bores.

I Kapittel 5 vil resultatene til eksperimentene i Kapittel 4 bli drøftet og tolket. På grunnlag av de resultatene som er funnet vil det i Kapittel 6 trekkes en konklusjon.

2 Beskrivelse av datasettet

I dette kapittelet blir loggene, som er brukt som grunnlag i dette prosjektet, presentert. Sensorene som finnes i de ulike loggene blir forklart, hva disse måler samt hva de målte verdiene $G_{c,d}$ kan brukes til. Videre blir de forskjellige litelogiene forklart, og hvordan de er presentert i en litelogiprofil. Til slutt vil det bli beskrevet en del potensielle problemer med loggene. Grunnen til dette er at problemene må tas hensyn til om det skal utvikles en loggtolker.

Dataene som er nødvendige for masteroppgaven kommer fra en boreplattform. På en borestreng sitter et visst antall sensorer som logger informasjon om formasjonen som det bores i. Ut i fra disse loggene skal det med høy sannsynlighet kunne sies hva slags litelogi det bores i.

Verdiene $G_{c,d}$ fra sensorene lagres i loggen med forskjellig enhet. Det som er fellesfaktoren for verdiene $G_{c,d}$ er at de er desimaltall som ligger i verdiområdet fra -100 til +20 000. Verdiene $G_{c,d}$ blir presentert i loggene i form av kolonner, hvor hver kolonne inneholder verdier $G_{c,d}$ målt fra en sensor. Verdiene $G_{c,d}$ fra sensorene blir målt med forskjellig frekvens fra logg til logg. I noen tilfeller logges verdiene $G_{c,d}$ hver gang borekronen har boret 0,1 meter og i andre tilfeller hver 0,5 meter.

En logg har et visst antall sensorer. En annen logg inneholder ikke nødvendigvis de samme, eller like mange sensorer. Grunnen til dette er at loggene kommer fra forskjellige boreplattformer hvor borestrengen kan være forskjellig.

Noe som er viktig å merke seg er at alle sensorene ikke vil ha målte verdier $G_{c,d}$ til enhver tid. Da vil det settes inn en ”dummyverdi” i loggen som er -999.25. Disse verdiene kan det ikke tas hensyn til når man klassifiserer.

Om sensorene egner seg til klassifisering, er avhengig av hvordan verdiene $G_{c,d}$ til hver enkelt litelogi er fordelt. Dersom fordelingen av verdiene $G_{c,d}$ til en litelogi ikke ligger i samme verdiområde som verdiene $G_{c,d}$ til en annen litelogi, vil sensoren egne seg godt til å skille disse to litelogiene. Skulle derimot disse to ligge i samme verdiområde må det brukes flere eller andre sensorer for å skille disse litelogiene fra hverandre. Mer informasjon om hvilke sensorer som blir logget og hvilke verdiområde de ligger i finnes i [11].

2.1 Beskrivelse av loggene som er brukt

Vi har tatt utgangspunkt i 11 forskjellige logger som det er mulig å benytte klassifiseringsmetodene på. Ut i fra disse loggene har vi gjort et utvalg av fire logger. Loggene er ment å fremstå som et representativt utvalg for alle loggene. Det har blitt valgt ut logger med forskjellig antall sensorer samt forskjellig litelogier å klassifisere til.

Den første loggen vi har blitt valgt ut vil bli kalt Logg 1 gjennom denne rapporten. Dette er en logg på rundt 1450 meter og registrerer verdier $G_{c,d}$ fra sensorene hver 0,1 meter. Loggen har fire forskjellige liteloger, Sandstone, Limestone, Claystone og Siltstone, og tillegg har den ni sensorer som kan brukes for å klassifisere litologiene, se Tabell 2.1 for mer informasjon.

Tabell 2.1: Sensorer i Logg 1

| Sensor | Beskrivelse |
|--------|--|
| 1 | Depth |
| 2 | High Resolution Bulk Density |
| 3 | High Resolution Gamma Ray (Mellomnivå) |
| 4 | Thermal Neutron Porosity (original Ratio Method) in Selected Lithology |
| 5 | Medium Resistivity / Mud Resistivity from HALS Measurements |
| 6 | Deep Resistivity |
| 7 | Medium Shallow Resistivity |
| 8 | Rate Of Penetration |
| 9 | High Resolution Gamma Ray |

Den andre loggen som har blitt brukt til å teste klassifiseringsmetodene blir kalt Logg 2. Denne loggen omfatter cirka 4500 meter og registrerer verdier $G_{c,d}$ hver 0,1 meter. Den inneholder to forskjellige liteloger, Sandstone og Claystone, og har 15 forskjellige sensorer. I Tabell 2.2 finnes en liste over hvilke sensorer som er med i denne loggen.

Tabell 2.2: Sensorer i Logg 2

| Sensor | Beskrivelse |
|--------|--|
| 1 | Depth |
| 2 | Bit Size |
| 3 | HRCC Caliper Calibrated |
| 4 | High Resolution Bulk Density |
| 5 | Delta Rho |
| 6 | High Resolution Gamma Ray |
| 7 | Thermal Neutron Porosity (original Ratio Method) in Selected Lithology |
| 8 | Extra Deep Resistivity (Mellomnivå) |
| 9 | Medium Resistivity / Mud Resistivity from HALS Measurements |
| 10 | Deep Resistivity |
| 11 | Deep Resistivity (Mellomnivå) |
| 12 | Medium Resistivity / Mud Resistivity from HALS Measurements (Mellomnivå) |
| 13 | Rate of Penetration |
| 14 | Shallow Resistivity |
| 15 | Medium Shallow Resistivity |

Den tredje loggen som vi har valgt å benytte blir kalt Logg 3. Den inneholder cirka 1600 meter med målinger \vec{G}_d , hvor sensorene blir registrert hver 0,1 meter. Loggen inneholder litelogiene Sandstone, Limestone og Claystone. Den inneholder seks forskjellige sensorer og de kan ses i Tabell 2.3.

Tabell 2.3: Sensorer i Logg 3

| Sensor | Beskrivelse |
|--------|--|
| 1 | Depth |
| 2 | High Resolution Bulk Density |
| 3 | High Resolution Gamma Ray |
| 4 | Thermal Neutron Porosity (original Ratio Method) in Selected Lithology |
| 5 | Medium Resistivity / Mud Resistivity from HALS Measurements |
| 6 | Deep Resistivity |

Den fjerde og siste loggen som klassifiseringsmetodene skal benyttes på blir kalt Logg 4. Loggen inneholder cirka 3000 meter med målinger \vec{G}_d . Disse blir målt hver 0,1524 meter og inneholder litelogiene Sandstone, Limestone, Shale og Anhydrite. Loggen inneholder ni sensorer, se Tabell 2.4.

Tabell 2.4: Sensorer i Logg 4

| Sensor | Beskrivelse |
|--------|--|
| 1 | Index |
| 2 | Depth |
| 3 | Delta-T / Sonic (also called Slowness or Interval Transit Time) |
| 4 | HRCC Caliper Calibrated |
| 5 | High Resolution Gamma Ray |
| 6 | High Resolution Bulk Density |
| 7 | Medium Resistivity / Mud Resistivity from HALS Measurements |
| 8 | Resistivity Shallow |
| 9 | Thermal Neutron Porosity (original Ratio Method) in Selected Lithology |

Det som er spesielt med Logg 4 er at det kun er tatt med de målingene \vec{G}_d hvor litelogiene er helt "rene", noe som vil si at det ikke er innslag av andre litelogier. Dette fører til at deler av loggen ikke er tolket til noen litelogi. Denne loggen var på formatet som en delvis tolket logg.

I de tre andre loggene har formasjonen blitt tolket ved hjelp av en litelogiprofil. Dette er en bildefil hvor man skisserer hvilken litelogi det bores i. Se Kapittel 2.3 for beskrivelse av en litelogiprofil [12].

2.2 Forklaring av sensorene

Her følger det en forklaring av de ulike sensorene som er nevnt i loggene over, samt deres betydning og hva de kan brukes til. Det vil også bli gitt noen begrunnelser på hvorfor ikke alle sensorene blir benyttet til å skille litelogiene. Se [10] for en mer detaljert forklaring av sensorene.

Index

Indekserer en linje i loggen, og egner seg ikke til klassifisering.

Depth Index

Denne sensoren indikerer dybden for borestrengen. Verdien G_{cd} til denne sensoren vil øke desto dypere det bores. Verdiene G_{cd} fra denne sensoren egner seg ikke til klassifisering.

Delta-T / Sonic (også kalt Slowness eller Interval Transit Time)

Delta-T / Sonic er en seismisk måling. Denne enheten sender lyd ut i formasjonen, og en mottaker registrerer hvor fort lyden beveger seg i de forskjellige formasjonene. Lyden vil forplante seg fortere i harde enn i løse eller ukonsoliderte formasjoner. Enheten for Delta-T / Sonic er mikrosekund per fot, μs / ft. Denne sensoren gir god informasjon om når det er et formasjonsskifte.

HRCC Caliper Calibrated

HRCC Caliper Calibrated er en måling som viser diameteren på hullet i tommer. Denne gir en indikasjon på om det er en fast formasjon eller om den har lett for å rase ut. Sensoren sier ikke direkte hva slags formasjon det bores i. Selve målingen blir utført ved hjelp av akustisk eller mekanisk teknologi som måler diameteren i hullet med fjærbelastede armer som dras mot hullveggen. Denne sensoren kan ikke benyttes til å skille litelogiene og er derfor utelatt.

Bit Size

Denne sensoren gir størrelsen på bittet eller boret. Den sier oss hva som er størrelsen på hullet som bores. Denne sammenliknes med sensoren HRCC Caliper Calibrated, og dersom HRCC Caliper Calibrated er større enn Bit Size betyr dette at hullet har rast ut. Sensoren Bit Size egner seg derfor ikke til å skille litelogiene.

High Resolution Gamma Ray

High Resolution Gamma Ray måler gammastråling fra formasjonen det bores i. Dette er en passiv måling, noe som vil si at den kun registrerer den gammastrålingen formasjonen gir fra seg. Sandstone har for eksempel lav gammastråling mens Shale har høy. High Resolution Gamma Ray brukes primært til å skille Shale fra de andre formasjonene.

High Resolution Bulk Density

High Resolution Bulk Density, eller romtetthet på norsk, måler tettheten til formasjonene. Enheten er gram per kubikkcentimeter, g/cm^3 . Dette er en aktiv måling, og det vil si at et radioaktivt materiale sender ut gammastråling. Mottakeren registrerer hvor mye av strålingen som mottas etter at den har vært i gjennom den formasjonen som bores i. En formasjon med høy tetthet vil gi lavere mottatt stråling enn i en formasjon med lav tetthet.

Medium Resistivity / Mud Resistivity from HALS Measurements

Resistivity, eller resistivitet på norsk, er en måling som viser strømmotstanden i en formasjon. Dette er en sensor som sender ut strøm i formasjonen på en gitt frekvens. Mottakeren registrerer en faseforskyvning og en demping på det utsendte signalet. Enheten er ohm meter, ohmm. Høy motstand kan indikere olje eller gass. Saltvann har lav motstand, og alle formasjoner offshore inneholder saltvann. Resistivitetsmålingen brukes primært til å skille olje og gass fra saltvann.

Resistivity Shallow

Resistivity Shallow er av samme type sensor som Medium Resistivity / Mud Resistivity from HALS Measurements, men "ser" ikke så dypt inn i formasjonen. Normalt sett er det alltid to eller flere resistivitetsmålinger med ulik inntrengningsevne. En Shallow Resistivity måling vil typisk "se" cirka 75 cm inn i formasjonen. Medium Resistivity / Mud Resistivity from HALS Measurements eller vanlig resistivitet trenger ca 1 – 1,3 meter inn i formasjonen. En Deep Resistivity måling vil typisk gå fra 2 – 3 meter inn i formasjonen. Siden inntrengningsevnen er så "dyp", brukes denne resistivitetsmålingen til geostyring i reservoaret.

Thermal Neutron Porosity (original Ratio Method) in Selected Lithology

Thermal Neutron Porosity er en måling som viser hvor stort tomrom (porøsitet) det er mellom kornene i en formasjon. Sandstone har høyere porøsitet enn en Shale. Dermed har Sandstone plass til mer væske mellom kornene, og vil derfor ha et større væskeinnhold. Gode typiske olje og gass reservoarer er Sandstone og Limestone, som begge har stor porøsitet. Enheten er m^3 / m^3 , og en verdi på 0,5 vil si 50 % hulrom i formasjonen mellom kornene. Thermal Neutron Porosity målingen er en aktiv måling, det vil si at et radioaktivt materiale sender ut nøytronstråling.

Rate of Penetration

Denne sensoren beskriver hvor rask borehastigheten er. Dersom verdien $G_{c,d}$ til denne sensoren er lav vil det si at det bores i en hard formasjon. Er verdien $G_{c,d}$ høy betyr dette at man borer i en løs formasjon. Denne sensoren kan ikke brukes til klassifisering.

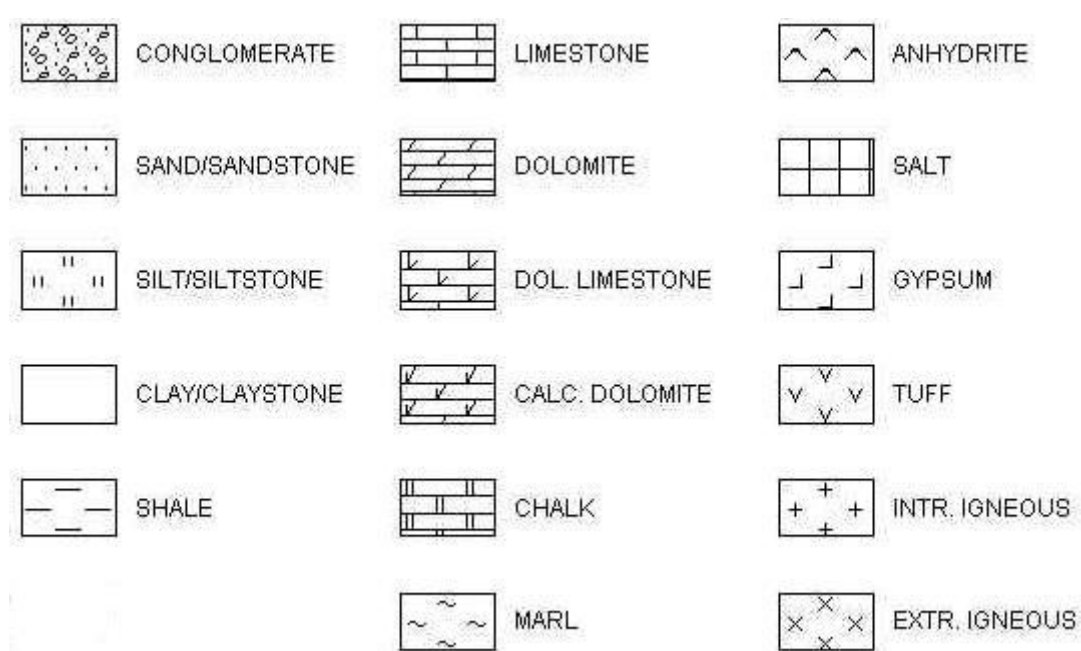
Delta Rho

Delta Rho er en sensor som brukes som en kvalitetsindikator på High Resolution Bulk Density målingen på hvor mye den er korrigert. Dersom Delta Rho har høye verdier $G_{c,d}$ vil dette si at målingen er mye korrigert, og ikke vil være til å stole på. Da må korrigeringen utforskes i større detalj. Dersom verdiene $G_{c,d}$ overskrider $\pm 0,15 \text{ g/cm}^3$ må dette gjøres. Denne sensoren egner seg ikke til klassifisering.

I tillegg finnes det sensorer som ligger på et mellomnivå mellom de andre sensorene. Disse er ikke tatt med når det klassifiseres.

2.3 Forklaring av litelogiene

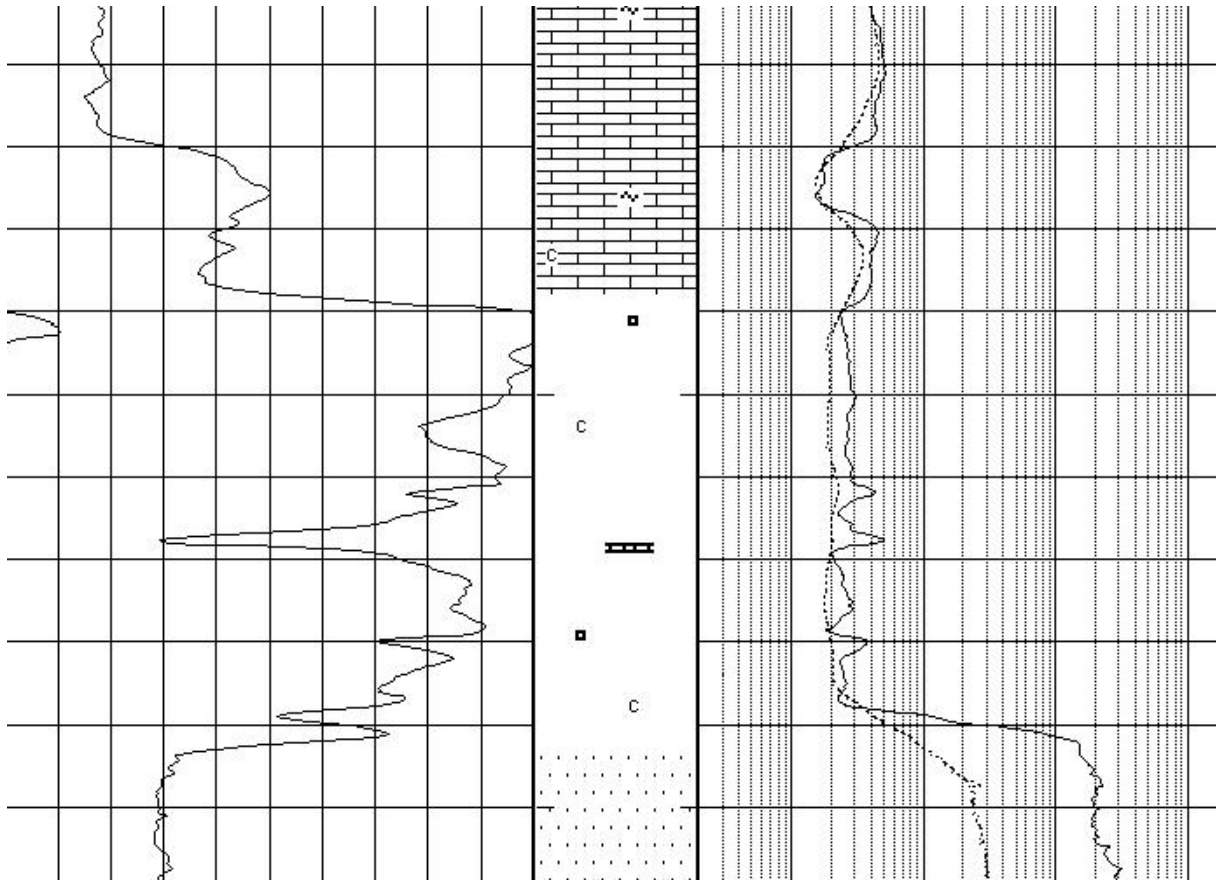
I loggene finnes det flere litelogier som beskriver hva det bores igjennom for å finne olje og gass. Litelogiene blir spesifisert av geologer som konstruerer en litelogiprofil som viser hvilke litelogier som finnes der det bores. Figur 2.1 illustrerer hvordan en geolog vil symbolisere de forskjellige litelogiene i en litelogiprofil. I tillegg til litelogiene i figuren kan også litelogiene Coal og Metamorphic forekomme.



Figur 2.1: Eksempel på hvordan litelogiene blir symbolisert i en logg

I de fire loggene som det har blitt tatt utgangspunkt i, er litelogiene Sandstone, Limestone, Shale, Anhydrite, Claystone og Siltstone representert.

Litelogiprofilen som en geolog lager kan man se et eksempel på i Figur 2.2. Denne figuren er et utsnitt fra Logg 3. Ved å benytte seg av symbolene i Figur 2.1 kan det ses av litelogiprofilen i Figur 2.2 at det først forekommer Limestone. Deretter går det over til litelogiene Claystone før det til slutt går over til Sandstone. Se [12] for en nærmere beskrivelse av hvordan en litelogiprofil kan bli laget og illustrert.



Figur 2.2: Eksempel på hvordan en logg ser ut

Grafen på venstre hånd i figuren viser hvordan High Resolution Gamma Ray varierer under boring, mens de andre grafen viser resistiviteten eller motstanden i litologien. Figur 2.2 viser hvordan High Resolution Gamma Ray forandrer seg når det bores i Claystone i forhold til Limestone og Sandstone. I tillegg vil resistiviteten gå opp når det bores i Sandstone. Dette betyr at Sandstone har større motstand enn litologiene Limestone og Claystone. Av figuren ser man at det er noen objekter i litologiene, og disse kalles sediment. Et sediment kan for eksempel være et fossil, som vil oppstå som et fremmedelement i litologiene. Se [10] for nærmere forklaring av sediment.

Det er denne litologiprofilen vi har benyttet når vi har laget de delvis tolkede loggene, hvor vi kobler målingene \vec{g}_d til riktig litologi. For å unngå at målingene \vec{g}_d blir klassifisert til feil litologi, noe som vil gå ut over kvaliteten på treningsgrunnlaget, ble det valgt å fjerne 1 – 2 meter i overgangene mellom litologiene. Dette fører til at litologiene ikke blir blandet og man får mer ”rene” litologier.

2.4 Potensielle problemer med loggene

Når loggene skal klassifiseres til riktig litologi er det en del faktorer i loggene som kan spille inn. Videre følger en innføring i problemene som er relatert til loggene, og problemene vil senere bli håndtert i Kapittel 4.

Den første faktoren som kan spille inn på klassifiseringsmetoden er at det ikke finnes like mange sensorer i de ulike loggene. Avhengig av hvilke sensorer som er med i en logg kan

klassifiseringssannsynligheten variere. Årsaken til dette er at noen sensorer kan medføre flere feil enn andre.

Når verdiområdet for en sensor forandrer seg kan det føre til problemer for klassifiseringsmetoden. Problemet vil være gjeldene hvis man skal klassifisere en måling \vec{G}_d fra en annen logg. Om verdiområdet forandrer seg vil det føre til at klassifiseringsmetoden ikke kan trene opp generelt. Med dette menes det at klassifiseringsmetodene må trenes opp med treningsdata fra samme borehull. Et eksempel på et verdiområde som kan variere, vil være at Claystone ikke utstråler den samme gammastråling på et sted i verden i forhold til et annet.

Verdiområdet kan forandre seg om det skulle inntreffe et shale-baseline-shift. Dette er et spesielt tilfelle som bare kan skje for sensoren High Resolution Gamma Ray, og gjør at gammaspekteret blir flyttet noen oktaver opp eller ned. Årsaken til dette er at boreslammet inneholder kalium eller at kalibreringen av sensoren High Resolution Gamma Ray ikke er helt rett. Et slikt shale-baseline-shift inntreffer ikke i noen av loggene som har blitt behandlet i dette prosjektet, og dermed har vi heller ikke hatt mulighet til å behandle dette problemet.

Et annet problem som kan oppstå er at loggene ikke inneholder de samme litelogiene. Det kan da bli vanskelig å lage en generell klassifiseringsmetode, dersom man ikke har treningsdata for alle sensorer til alle litelogier. Har man ikke treningsgrunnlag for en litologi, kan man heller ikke klassifisere en måling \vec{G}_d til denne litologien.

En annen faktor som vil gjøre det vanskelig for klassifiseringsmetoden, er at sensorene ikke vil ha registrerte verdier $G_{e,d}$ til enhver tid. Det logges da såkalte "dummyverdier", og dette vil kunne føre til at feilmålinger får stor innflytelse under trening. Årsaken til dette er at det brukes middelværdi og standardavvik som treningsgrunnlag, og finnes det lite treningsdata kan feilmålinger føre til store endringer i disse. Se Kapittel 3 for forklaring av middelværdi og standardavvik.

Sedimenter som gjør innslag i litelogiene er også et mulig problem som må tas hensyn til når litelogiene skal klassifiseres. Årsaken er at sedimenter gjør utslag, spesielt på resistivitetmålingene, ved at verdiene $G_{e,d}$ blir høye. Verdiområdet til for eksempel High Resolution Gamma Ray blir da mye større enn den ville vært dersom sedimentene ikke hadde vært til stede. Se Figur 2.2 for et eksempel på dette.

Den samme sensoren kan bli registrert på forskjellige måter i forskjellige logger. Et eksempel kan være om sensoren Thermal Neutron Porosity blir logget i prosent i en logg i forhold til en annen. Årsaken til dette er at det er forskjellige leverandører av utstyret til boreplattformer. Dette kan føre til feil i klassifiseringsmetodene om dette ikke blir tatt hensyn til.

Da tolkningen av litelogiprofilen er gjort av oss, er det usikkert om treningsgrunnlaget er så godt som det burde være. Om en geolog hadde skrevet ned hvilke litelogier som forekommer hvor i loggene, kunne dette gjort klassifiseringen sikrere. Vi har ikke tilstrekkelig kunnskap for å tolke en litelogiprofil, og feil i tolkningen kan forekomme. Litelogiprofilene er presentert på et format der dybden blir markert hver femte meter. I noen logger ble det gjort målinger \vec{G}_d hver 0,1 meter, og det kan føre til noen unøyaktigheter. Mange skift mellom litelogiene gjør at det kan forekomme enkelte menneskelige feil.

3 Bayesiansk beslutningsteori

I dette kapitlet vil metodene som vi har brukt i dette prosjektet bli forklart. Det blir også gitt en begrunnelse på hvorfor de har blitt valgt, samt deres forutsetninger.

En mulig tilnærming til problemet som omhandler automatisk tolkning av logger, er å benytte seg av Bayesiansk beslutningsteori [2][3]. Denne tilnærmingen brukes ofte for å løse problemer vedrørende mønstergjenkjenning. Her vil det bli vist hvordan teorien kan benyttes som løsning på hovedproblemstillingen, se Kapittel 1.2.

Når logger som blir registrert fra en boretplattform skal tolkes, kan maskinlæring benyttes. Dette er et felt som blant annet brukes for å klassifisere data inn i grupper. Her blir det tatt utgangspunkt i ledet læring. Dette består av å lese inne ferdig analysert data slik at man i etterkant kan klassifisere nye ukjente målinger.

Maskinlæring er en metode som kan brukes for å trene opp en datamaskin til å gjøre selvstendige valg, uten hjelp fra et menneske. Ved klassifiseringsprosessen blir en maskin trent opp ved at den blir ledet til å se sammenhengene mellom målinger og dens tilhørende tilstand. Årsaken til at maskinlæring har blitt valgt er at det foreligger eksempeldata slik at det gjør det mulig å trene opp datamaskinen til å klassifisere litelogiene.

I dette kapitlet gir vi først en innføring i Bayes Teorem [1]. Deretter vil vi vise hvordan normalfordeling kan benyttes til å klassifisere litelogiene ved hjelp av én sensor. Videre viser vi hvordan Naive Bayes Classifier [4], som kan utledes fra Bayes Teorem, kan benyttes til å klassifiserer litelogiene ved hjelp av flere sensorer. Til slutt vil problemet utvides ved å ta hensyn til før og etterliggende målinger \vec{g}_a . Dette gjør vi ved å utvide Naive Bayes Classifier med Hidden Markov Model [5][6].

3.1 Bayes Teorem

Vi har valgt å benytte metoder innen Bayesiansk beslutningsteori til å skille litelogiene. I dette delkapitlet gir vi en innføring i Bayes Teorem. Årsaken til at vi valgte denne beslutningsteorien er at den er enkel å implementere, samtidig som den gir en gradering av hvor sikker klassifiseringen er. Bayes Teorem er enkel å oppdatere slik at den kan benyttes på mer konkrete problemstillinger.

Vi lar $\{\omega_1, \omega_2, \dots, \omega_j\}$ være et bestemt antall tilstander j , og vi lar $\{\alpha_1, \alpha_2, \dots, \alpha_i\}$ være et bestemt antall utfall i . Sannsynligheten $\lambda(\alpha_i | \omega_j)$ beskriver tapet ved å velge utfall α_i når tilstanden ω_j er den riktige, x er en d -komponent vektor variabel. Videre lar vi $p(x | \omega_j)$ være den tilstandsavhengige tetthetssannsynlighetsfunksjonen for x , hvor tetthetssannsynlighetsfunksjonen for x , betinget av ω_j er den riktige tilstanden. $P(\omega_j)$ beskriver den a priori sannsynligheten til tilstanden ω_j . Ut i fra dette kan den nye sannsynligheten $P(\omega_j | x)$ bli anslått ved hjelp av $p(x | \omega_j)$ ved å bruke Bayes Teorem [1]:

$$P(\omega_j | x) = \frac{p(x | \omega_j)P(\omega_j)}{p(x)}$$

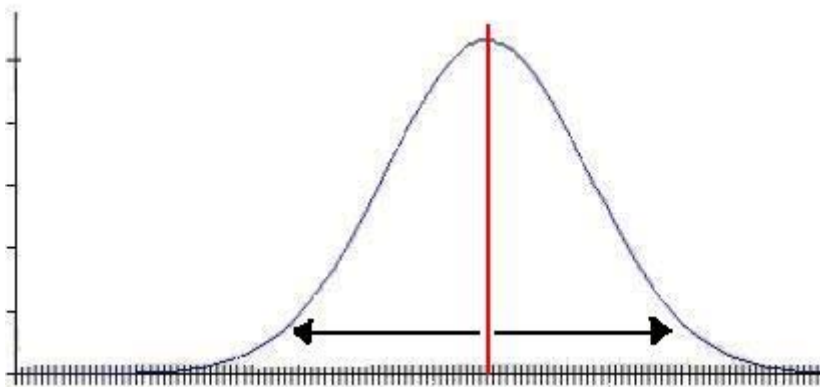
3.2 Normalfordeling

Bayes Teorem består av den betingede sannsynligheten $p(x|\omega_j)$ og den a priori sannsynligheten $P(\omega_j)$. Dette teoremet kan også benyttes hvor vektor variabelen x for en gitt tilstand ω_j er kontinuerlig fordelt. Se [1] for nærmere forklaring av denne situasjonen.

Hvis vi har en observasjon x kan $p(x|\omega_j)$ angis ved formel for normalfordeling.

$$p(x|\omega_j) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Ut i fra formelen kan man se at normalfordelingen er spesifisert av to parametere, middelveidien μ og standardavviket σ . Ofte blir dette betegnet som $p(x) \sim N(\mu, \sigma)$ for å si at x er normalfordelt med middelveidien μ og standardavviket σ . Se avsnitt 3.3 og 3.4 for forklaring av henholdsvis middelveid og standardavvik.



Figur 3.1: Eksempel på en normalfordeling

I Figur 3.1 viser et eksempel på en normalfordeling der den loddrette streken indikerer middelveidien μ og hver av de vertikale pilene indikerer standardavviket σ .

3.3 Middelerverdi

Estimert middelerverdi er gjennomsnittet av et endelig antall observasjoner. Observasjonene i en normalfordeling har en tendens til å samle seg rundt en middelerverdi. Hensikten med å finne middelerverdien i denne sammenheng er at den benyttes til å regne ut standardavviket, samt at den brukes i formelen for normalfordeling.

Formel for estimert middelerverdi er:

$$\hat{\mu}_x = \frac{\sum_{i=0}^n X_i}{n}$$

I formelen står x_i for en enkelt observasjon, og n er antall observasjoner totalt. Når middelerverdi blir brukt videre i rapporten menes estimert middelerverdi.

Eksempel 3.1

Her innføres det et eksempel for å kunne beskrive hvordan middelerverdien beregnes. En tenkt litelogi Sand, har følgende verdier $G_{c,d}$: 1, 2, 3, 3, 4, 5. Middelerverdi for denne litelogien blir da som følger:

$$\hat{\mu} = \frac{1+2+3+3+4+5}{6} = 3.$$

Middelerverdien for denne fordelingen er 3.

3.4 Standardavvik

Estimert standardavvik til en fordeling er et estimat som forteller i hvilken grad verdiene fra fordelingen kan forventes å variere. Denne variansen forteller hvordan observasjonene forventes å fordele seg rundt middelerverdien. Grunnen til innføringen av standardavvik er at denne verdien er sentral i normalfordelingen som er forklart i Kapittel 3.2.

Formel for estimert standardavvik er:

$$\hat{\sigma}_x = \sqrt{\frac{\sum_{i=0}^n (x_i - \mu)^2}{n-1}}$$

I formelen står x_i for en enkelt observasjon, μ er middelerverdien og n er antall observasjoner. Når uttrykket standardavvik blir brukt videre i rapporten menes estimert standardavvik.

Eksempel 3.2

Dette er en fortsettelse av Eksempel 3.1, og de samme verdiene $G_{c,d}$ som ble innført i det eksempelet blir benyttet for å beskrive hvordan standardavviket blir utregnet. Beregningen av standardavviket blir som følger:

$$\hat{\sigma} = \sqrt{\frac{(1-3)^2 + (2-3)^2 + (3-3)^2 + (3-3)^2 + (4-3)^2 + (5-3)^2}{6-1}} = 1,4142$$

Standardavviket for denne fordelingen blir 1,4142.

3.5 Klassifisering ved hjelp av én sensor

For å klassifisere litelogene ved hjelp av én sensor, trengs det en klassifiseringsmetode. Vi antar at observasjonene er normalfordelte med kontinuerlige målinger. Derfor har vi valgt å bruke formelen for normalfordeling til å klassifisere litelogene.

Ved klassifisering av en observasjon, må det bestemmes hvor sannsynlig det er at den tilhører en tilstand. Om det finnes fire tilstander må det beregnes fire sannsynligheter. Den tilstanden som får høyest sannsynlighet er da den tilstanden som observasjonen mest sannsynlig tilhører.

Normalfordeling gir en gradering av hvor sannsynlig det er at en observasjon tilhører en tilstand som et tall mellom 0 og ∞ . Dette betyr at desto høyere tall man får, desto mer sikker er den.

Det forutsettes at utregningene av middelværdi og standardavvik, henholdsvis fra Kapittel 3.3 og 3.4, er utført for å kunne beregne sannsynlighetene som normalfordelingen anslår. Disse utregningene må gjøres for hver tilstand.

Eksempel 3.3

Her følger et eksempel på hvordan formelen for normalfordeling blir brukt. Verdiene fra Eksempel 3.1 og 3.2 videreføres for å vise hvordan en klassifisering blir gjort. I tillegg må det også innføres en ny litologi. Denne litologien, Stein, har følgende verdier $G_{c,d}$: 11, 12, 13, 13, 14, 15.

Det må beregnes middelværdi og standardavvik for litologien Stein. Ved bruk av formlene fra Kapittel 3.3 og 3.4, blir middelværdi 13 og standardavviket 1,4142. Om den registrerte verdien $G_{c,d}$ er 15, og det er ønskelig å finne ut hvilken litologi denne verdien $G_{c,d}$ mest sannsynlig tilhører, brukes formel for normalfordeling på følgende måte:

Litologien Sand:

$$f(x) = \frac{1}{1,4142\sqrt{2\pi}} \exp\left(-\frac{(15-3)^2}{2 * 1,4142^2}\right) = 6,53879 * 10^{-17}$$

Litologien Stein:

$$f(x) = \frac{1}{1,4142\sqrt{2\pi}} \exp\left(-\frac{(15-13)^2}{2 * 1,4142^2}\right) = 0,10378$$

Verdien $G_{c,d}$ 15 tilhører mest sannsynlig litologien Stein, fordi formelen for normalfordeling gir høyest verdi for denne litologien.

3.6 Naive Bayes Classifier

I dette delkapittelet vil det statistiske verktøyet Naive Bayes Classifier [4] som er basert på Bayes Teorem bli tatt for seg. Denne metoden er spesielt egnet når det skal klassifiseres ved hjelp av flere sensorer. Naive Bayes Classifier kan utledes fra Bayes Teorem når observasjonene er uavhengig av hverandre. Se [1] for utledning av forholdet mellom Bayes Teorem og Naive Bayes Classifier.

Naive Bayes Classifier velger ut den mest sannsynlige klassifikasjonen V_{nb} , gitt observasjonene $X = x_1, x_2, x_3, \dots, x_n$. X tilsvarer \vec{G}_d i denne metoden. Det forsøkes å lære $f : X \rightarrow V$, hvor $V = \{v_1, v_2, v_3 \dots v_t\}$ er tilstandene en verdi $G_{c,d}$ blir klassifisert til. Det vil bli dannet treningspar $\langle x_i, v_i \rangle$ som vil representere datasettet.

Den generelle formelen for Naive Bayes Classifier er:

$$V_{nb} = \arg \max_{v_j \in V} P(v_j) \prod p(x_i | v_j)$$

Ved hjelp av treningssettet blir $\prod p(x_i | v_j)$ brukt til å beregne sannsynlighetene for at observasjonene tilhører tilstandene. Sannsynligheten $P(v_j)$ beskriver hvor stor muligheten for at nettopp den spesifikke tilstanden forekommer i treningssettet.

3.6.1 Klassifisering av flere sensorer

Når man skal klassifisere flere sensorer, benyttes Naive Bayes Classifier. Dette gjøres ved at man benytter seg av formelen for normalfordeling. Formelen brukes da til å regne ut $P(x_i | v_j)$, slik som formelen under viser:

$$p(x_i | v_j) = \frac{1}{\sigma_{v_j} \sqrt{2\pi}} \exp\left(-\frac{(x_i - \mu_{v_j})^2}{2\sigma_{v_j}^2}\right)$$

Likningen over viser sammenhengen mellom en betinget sannsynlighet og en normalfordeling. Sammenhengen kommer frem i formelen for normalfordeling ved at man benytter seg av middelværdien og standardavviket til dens tilhørende tilstand. Se [1] for utledningen av denne sammenhengen.

Dersom man sammenlikner formelen for klassifisering med formelen for Naive Bayes Classifier ser man at $P(v_j)$ er utelatt. Grunnen til dette er at det blir antatt at det er like stor sannsynlighet for at alle litelogiene kan forekomme. Dermed utgjør den et like stort utslag i alle litelogiene, og kan dermed utelates da en ikke har noen funksjon.

Eksempel 3.4

I dette eksempelet blir det vist hvordan Naive Bayes Classifier blir brukt ved hjelp av formel for normalfordeling. Her benyttes verdiene som ble utregnet i Eksempel 3.3, men det innføres i tillegg en ny sensor for å kunne vise hvordan det klassifiseres ved hjelp av flere sensorer.

Det innføres derfor en ny middelværdi og et nytt standardavvik for litelogene til Sensor 2. Her settes $\mu_{Sand} = 2$, $\sigma_{Sand} = 1$, $\mu_{Stein} = 3$ og $\sigma_{Stein} = 1$. Den nye verdien $G_{c,d}$ som blir målt er 3, og utregningen av Sensor 2 blir som følger:

Litelogien Sand, Sensor 2:

$$f(x) = \frac{1}{1\sqrt{2\pi}} \exp\left(-\frac{(3-2)^2}{2*1^2}\right) = 0,24197$$

Litelogien Stein, Sensor 2:

$$f(x) = \frac{1}{1\sqrt{2\pi}} \exp\left(-\frac{(3-3)^2}{2*1^2}\right) = 0,39894$$

Dette må gjøres for alle sensorene som klassifiseringsmetoden skal ta hensyn til. I eksempelet er det kun blitt tatt hensyn til to sensorer for å vise hvordan formelen fungerer. Videre må alle sensorene multipliseres sammen. Med dette menes det at alle resultatene fra normalfordeling, som blir beregnet ut i fra samme litelogi, multipliseres sammen. I eksempelet over blir resultatet som følger:

Litelogien Sand:

$$(6,54 * 10^{-17}) * 0,2419 = 1,58 * 10^{-17}$$

Litelogien Stein:

$$0,1038 * 0,3989 = 0,0414$$

Ut fra resultatet ser vi at metoden angir at det er mest sannsynlig at det bores i litelogien Stein fordi den har høyest verdi. En viktig ting å merke seg er at tallene fort kan bli veldig små. For at tallene skal blir lettere å behandle, kan det lønne seg å beregne logaritmen.

3.7 Hidden Markov Model

Litelogiprofilen viste at når en litelogi forekom på en dybde var det stor sjanse for at målingen \vec{G}_d på neste dybde også var den samme litelogien. Vi valgte derfor å undersøke om det fantes en korrelasjon mellom etterfølgende målinger \vec{G}_d , og Hidden Markov Model ble valgt til å undersøke om det fantes en slik sammenheng. Årsaken til valget av denne metoden var at den bygger på Naive Bayes Classifier, og har derfor de samme fordelene. I tillegg benytter den en eventuell korrelasjon til å forbedre klassifiseringen.

Hidden Markov Model [5][6] er et kraftig verktøy for å kunne utføre statistiske analyser på en gitt tallrekke. En Hidden Markov Model kan angis ved hjelp av tre parametere, og for enkelthets skyld brukes den kompakte notasjonen $\lambda = (A, B, \pi)$. Parametrene defineres på følgende måte:

$$A = [a_{ij}], a_{ij} = P(q_{t+1} = S_j | q_t = S_i), \quad 1 \leq i, j \leq N$$

$$B = [b_j(k)], b_j(k) = P(x_t = v_k \text{ tid } t | q_t = S_j), \quad 1 \leq j \leq N, 1 \leq k \leq M$$

$$\pi = [\pi_i], \pi_i = P(q_1 = S_i), \quad 1 \leq i \leq N$$

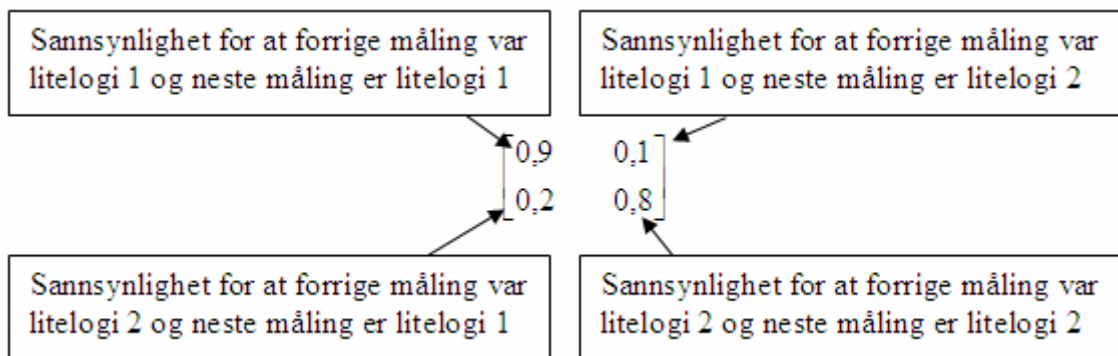
Utledningen av disse parametrene følger i avsnittene under. For å kunne definere de tre parametrene i formelen må vi først formulere et tilstandsalfabet, S , med N antall tilstander. Tilstandsalfabetet kan angis som $S = s_1, s_2, s_3, \dots, s_N$, og tilstanden ved tidspunkt t angis som q_t . V er symbolalfabetet hvor M er antall distinkte observasjoner per tilstand, det vil si antall alfabetstørrelsen. Symbolalfabetet kan angis som $V = v_1, v_2, v_3, \dots, v_M$.

Gitt passende verdier for N, M, A, B og π kan Hidden Markov Model bli brukt som en generator for å angi en observasjonssekvens:

$$O = O_1, O_2, O_3, \dots, O_T$$

hvor hver observasjon O_t er et av symbolene fra V , og T er antall observasjoner i sekvensen. Se [5] for sammenhengen mellom verdiene N, M, A, B, π og observasjonssekvensen O .

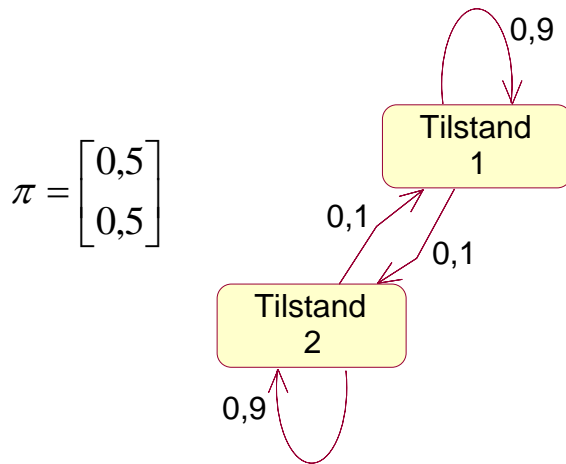
Når disse er definert kan det videre defineres parameteren A i Hidden Markov Model. A er en overgangsmatrise som inneholder sannsynlighetene for at tilstand j blir etterfulgt av tilstand i . Beskrevet på en annen måte forteller den hva som er sannsynligheten for å gå fra en tilstand til hvilken som helst annen i ett skritt, eller forbli i den samme tilstanden. I en stasjonær modell vil matrisen være konstant. Se Figur 3.2 for oppbygning av overgangsmatrisen.



Figur 3.2: Eksempel på en overgangsmatrise

Parameteren B i Hidden Markov Model beskriver en såkalt observasjonsmatrise. Denne matrisen holder sannsynlighetene for observasjonen k for tilstand j , uavhengig av t . Beskrevet på en annen måte så er B symbolsannsynlighetsdistribusjonen i tilstand j .

Den siste parameteren i formelen for Hidden Markov Model er π . Denne matrisen inneholder initialsannsynlighetene. Matrisen vil kun bli tatt hensyn til i oppstarten av Hidden Markov Model, noe også navnet på matrisen tilsier. Se Figur 3.3 for visualisering.



Figur 3.3: Eksempel på initialmatrise og overgangssannsynlighet

Initialmatrisen i figuren tilsier at det er like stor sannsynlighet for å gå til de to tilstandene. I tilstand 1 og 2 vises det hvor stor sannsynlighet det er for å gå til motsatt tilstand, samt hvor stor sannsynlighet det er for å være i bli værende i samme tilstand.

Med Hidden Markov Model ønsker vi å regne ut sannsynligheten for observasjonsrekken $O = O_1, O_2, \dots, O_T$ gitt modellen λ , $P(O | \lambda)$. Den enkleste måten å gjøre dette på er å telle opp antall mulige tilstandsrekkefølger av lengde T , altså antall observasjoner. Dersom vi antar at vi har en gitt tilstandsrekkefølge, $Q = q_1, q_2, q_3, \dots, q_T$. Kan sannsynligheten for observasjonsrekken O gitt modellen λ , $P(O | \lambda)$, utledes til å få formel under. Se [5] for full utledning av formel og forklaring av symboler.

$$\begin{aligned} P(O | \lambda) &= \sum_Q P(O | Q, \lambda) P(Q | \lambda) = \sum_Q P(O, Q | \lambda) \\ &= \sum_{q_1, q_2, \dots, q_T} \pi_{q_1} b_{q_2}(O_1) a_{q_1 q_2} b_{q_2}(O_2) \dots a_{q_{T-1} q_T} b_{q_T}(O_T) \end{aligned}$$

Formelen vil ikke bli utledet og årsaken til at denne er tatt med er for å vise hvor kompleks metoden over er. Den er dessverre veldig ineffektiv, fordi den fører til så mange beregninger av Hidden Markov Model. Selv for små verdier av $N = 2$ (tilstander) og $T = 200$ (observasjoner) vil dette føre til 10^{60} beregninger om metoden over benyttes.

På grunn av det store antallet beregninger, er det ønskelig å benytte en annen fremgangsmåte. Forward – Backward prosedyren [5] benyttes da denne ikke fører til like mange beregninger.

3.8 Forward – Backward Prosedyren

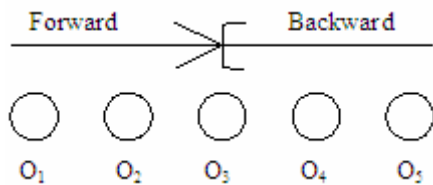
Forward – Backward prosedyren går ut på å beregne sannsynlighetene fra starten O_1 til gjeldene tilstand O_t , dette kalles forward. Deretter beregnes alle sannsynlighetene fra slutten O_T til gjeldene tilstand O_{t+1} , og denne blir kalt backward. Prosedyren gjentas for alle gitte observasjoner. Forward – Backward prosedyren angis ved å multiplisere sammen forward variabelen, $\alpha_t(i)$ og backward variabelen $\beta_t(i)$ på følgende måte:

$$\alpha_t(i) * \beta_t(i) = P(O | \lambda)$$

hvor

$$\alpha_t(i) = P(O_1 O_2 \dots O_t, q_t = S_i | \lambda)$$

$$\beta_t(i) = P(O_{t+1} O_{t+2} \dots O_T | q_t = S_i, \lambda)$$



Figur 3.4: Eksempel på Forward – Backward

I Figur 3.4 ser man et eksempel på hvordan forward og backward regnes ut i observasjon O_3 . Forward omfatter alle observasjonene fra starten til O_3 , mens backward omfatter alle observasjonene fra slutten til og med observasjon O_3 .

3.8.1 Forward algoritmen

Forward algoritmen regner ut sannsynligheten for hver tilstand S , gitt tidligere observasjoner $O_1 O_2 \dots O_t$. Forward variabelen $\alpha_t(i)$ blir som følger:

$$\alpha_t(i) = P(O_1 O_2 \dots O_t, q_t = S_i | \lambda)$$

Gitt at vi har modellen λ , og sannsynlighetene til den partielle sekvensen $O_1 O_2 \dots O_t$ og tilstand S_i på tidspunktet t , kan $\alpha_t(i)$ løses slik:

1. Initialisering:

$$\alpha_1(i) = \pi_i b_i(O_1), \quad 1 \leq i \leq N$$

2. Induksjon:

$$\alpha_{t+1}(j) = \left[\sum_{i=1}^N \alpha_t(i) a_{ij} \right] b_j(O_{t+1}), \quad 1 \leq t \leq T-1, \quad 1 \leq j \leq N$$

3. Terminering:

$$P(O | \lambda) = \sum_{i=1}^N \alpha_T(i)$$

Trinn 1 initierer forward sannsynlighetene som en kombinasjon av tilstand S_i og observasjon O_1 . Her er alle verdiene i initialmatrisen, π_i , satt til $1/N$, hvor N er antall tilstander. Trinn 2, som er kjernen i forward algoritmen, sier hvordan en tilstand S_j kan bli nådd ved tidspunktet $t + 1$ fra N mulige tilstander S_i , $1 \leq i \leq N$, ved tidspunktet t .

Siden $\alpha_t(i)$ er sannsynligheten for at kombinasjonen av $O_1 O_2 \dots O_t$ er observert og at tilstanden ved tidspunkt t er S_i . Da er produktet $\alpha_t(i) a_{ij}$ sannsynligheten for at kombinasjonen av $O_1 O_2 \dots O_t$ er observert, og at tilstand S_j er nådd ved tidspunktet $t + 1$ via tilstand S_i på tidspunktet t . Om produktet summeres over alle mulige tilstander N på tidspunktet t , resulterer det i sannsynligheten for S_j ved tidspunktet $t + 1$ med alle de tidligere partielle observasjonene som er ledsagende.

Når S_j er kjent beregner man $\alpha_{t+1}(j)$ ved å redegjøre for observasjon O_{t+1} i tilstand j . For eksempel kan dette gjøres ved å multiplisere med den summerte mengden av sannsynligheten $b_j(O_{t+1})$. Utregningen av induksjonen blir gjort for alle tilstander j , $1 \leq j \leq N$ og blir iterert for alle $t = 1, 2, \dots, T - 1$. Trinn 3 regner til slutt ut $P(O | \lambda)$ som summen av den avsluttende forward variabelen $\alpha_T(i)$.

Dersom man sammenlikner antall beregninger som må gjøres for Hidden Markov Model når man regner ut $\alpha_t(i)$, og man har $N = 2$ (tilstander) og $T = 200$ (observasjoner). Da vil det i dette tilfellet føre til kun 800 beregninger i motsetning til 10^{60} .

3.8.2 Backward algoritmen

Backward algoritmen regner ut sannsynligheten for hver tilstand S gitt etterliggende observasjoner $O_{t+1} O_{t+2} \dots O_T$. Dette er det motsatte av forward, der det ses på alle tidligere observasjoner. Backward variabelen $\beta_t(i)$ blir gitt på følgende måte:

$$\beta_t(i) = P(O_{t+1} O_{t+2} \dots O_T | q_t = S_i, \lambda)$$

På bakgrunn av sannsynligheten til den partielle sekvensen fra $t + 1$ til slutten, gitt tilstand S_i , på tidspunktet t og modellen λ , kan $\beta_t(i)$ løses slik.

1. Initialisering:

$$\beta_T(i) = 1, \quad 1 \leq i \leq N$$

2. Induksjon:

$$\beta_t(i) = \sum_{j=1}^N a_{ij} b_j(O_{t+1}) \beta_{t+1}(j), \quad t = T - 1, T - 2, \dots, 1, \quad 1 \leq i \leq N$$

Trinn 1 definerer alle verdier av $\beta_T(i)$ til å være 1 for alle i , dette vil si at alle verdier i initialmatrisen blir satt til 1. Induksjonen i trinn 2 sier at for å ha vært i tilstand S_i ved tidspunkt t , må det gjøres rede for observasjonssekvensen ved fra tidspunkt $t + 1$ og så videre. Det må tas i betraktning alle mulige tilstander for S_j ved tidspunkt $t + 1$ gitt ved overgangen fra S_i til S_j , betegnelsen a_{ij} , og observasjonen O_{t+1} i tilstand j , betegnelsen $b_j(O_{t+1})$. Til slutt

må det tas med i beregningen den partielle observasjonssekvensen fra tilstand j , betegnelsen $\beta_{t+1}(j)$. Her kreves det like mange beregninger som ble utført i forward for å regne ut $\beta_t(i)$.

Eksempel 3.5

En videreføring av Eksempel 3.4 viser hvordan Hidden Markov Model fungerer i praksis. Først vil det bli vist hvordan forward algoritmen blir brukt av Hidden Markov Model. Det må settes opp en initialmatrise som settes til $1/N$, og i dette tilfellet blir den $1/2$.

Videre må det lages en overgangsmatrise. Det kan gjøres ved å manuelt telle opp antall skift fra en litelogi til enten seg selv eller andre litelogier. Så regnes det ut sannsynligheten for at et skifte vil inntreffe. I overgangsmatrisen settes overgangssannsynlighetene opp etter et fast mønster, som vist i Figur 3.2.

Videre i beregningene antas det at overgangsmatrisen er den samme som man finner igjen i denne figuren. Videre benyttes matriseverdiene som vist under:

Litelogien Sand:
 $(0,5 * 0,9) + (0,5 * 0,2) = 0,55$

Litelogien Stein:
 $(0,5 * 0,1) + (0,5 * 0,8) = 0,45$

Ved å multiplisere svaret fra beregningene over med verdiene man fikk fra Naive Bayes Classifier får man en ny matrise. Når neste observasjon skal klassifiseres, blir denne matrisen brukt på samme måte som initialmatrisen blir benyttet i beregningene over. Beregningen av denne matrisen blir vist under:

$$0,55 * (1,58 * 10^{-17}) = 8,69 * 10^{-18}$$

$$0,45 * 0,0414 = 0,02863$$

Backward blir beregnet på omtrent den samme måten som forward, men med en annen initialmatrise samt at overgangsmatrisen blir brukt på en annen måte. Initialmatrisen blir satt til 1 for alle litelogier. Beregningen av backward blir utført på samme observasjon som forwardverdien.

Litelogien Sand:
 $(1 * 0,9) + (1 * 0,1) = 1$

Litelogien Stein:
 $(1 * 0,2) + (1 * 0,8) = 1$

Verdiene fra Naive Bayes Classifier multipliseres inn på samme måte som over. Observasjonsmatrisen for backward blir da:

$$1 * (1,58 * 10^{-17}) = 1,58 * 10^{-17}$$

$$1 * 0,0414 = 0,0414$$

Ved å multiplisere forward og backward, som vist i Figur 3.4, fås det hvilken litelogi Hidden Markov Model foreslår. I dette eksempelet blir svaret:

$$0,55 * (1,58 * 10^{-17}) = 8,69 * 10^{-18}$$

$$0,45 * 0,0414 = 0,02863$$

For at Hidden Markov Model skal fungere trengs det flere observasjoner over tid enn den ene som er brukt i eksemplet over. Eksempelet viser kun hvordan beregningene blir foretatt i Hidden Markov Model, ikke hvordan Hidden Markov Model fungerer i praksis.

4 Eksperimenter

I dette kapittelet blir det beskrevet de eksperimentene som vi har utført i løpet av prosjektperioden. Metoden leave-one-out, som brukes til å trene opp systemet best mulig når man har begrenset tilgang til treningsdata, blir forklart. Klassifiseringen ved hjelp av én sensor vil bli beskrevet, og deretter Naive Bayes Classifier som benytter flere. Så beskrives det hvordan Hidden Markov Model kan benyttes til klassifisering av litelogiene på bakgrunn av før og etterliggende målinger \vec{G}_a . Det har også blitt utført optimaliseringer av Naive Bayes Classifier og Hidden Markov Model for å utbedre klassifiseringen. Undersøkelser av hvor godt litelogiene klassifiseres dersom klassifiseringsmetodene trenes opp ved hjelp av en logg for deretter å klassifisere målinger \vec{G}_a fra en annen, er også blitt gjennomført.

4.1 Leave-one-out

Leave-one-out metoden blir brukt for å gi et best mulig estimat på feilklassifiseringsraten når kun et begrenset antall målinger \vec{G}_a , kan brukes til trening og testing.

Metoden går ut på å ta utgangspunkt i alle målingene \vec{G}_a som er klassifisert til en litelogi, og ut i fra mengden trekkes det ut en måling \vec{G}_a . Resten blir brukt som treningsgrunnlag. Videre blir den utvalgte målingen \vec{G}_a forsøkt klassifisert til en litelogi, og deretter kontrolleres det om den klassifiserte litelogien er den samme som den virkelige litelogien. Om den er lik, har målingen \vec{G}_a blitt klassifisert riktig, ellers er den klassifisert feil. Dette gjentas til alle målingene \vec{G}_a har blitt klassifisert én gang.

Resultatene fra leave-one-out metoden blir presentert i en forvirringsmatrise. Denne matrisen er en fremstilling av antall målinger \vec{G}_a som blir klassifisert riktig. Hvis en måling \vec{G}_a blir feilklassifisert vil matrisen fortelle hvilken litelogi den blir feilklassifisert til. Målinger \vec{G}_a som blir klassifisert riktig vil ligge diagonalt, fra venstre mot høyre i forvirringsmatrisen. Se Figur 4.1 for et eksempel på en slik forvirringsmatrise.

Eksempel 4.1

Dette eksempelet bygger på verdiene fra Eksempel 3.3, og viser bruken av leave-one-out metoden. Vi velger ut verdien $G_{c,d}$ 15 fra mengden til litelogien Stein, så blir middelverdien og standardavviket for Stein uten denne verdien $G_{c,d}$ beregnet. Formlene fra Kapittel 3.3. og 3.4 blir benyttet og vi finner at middelverdien blir 12,6 og standardavviket blir 1,16. Middelverdi og standardavvik for Sand er det samme.

Vi forsøker å klassifisere verdien $G_{c,d}$ 15, og dette gjøres ved å benytte formelen fra Kapittel 3.5. Beregningen av fører at resultatet for litelogien Stein blir 0,0405. Ved å gjøre det samme for litelogien Sand, fås det samme resultatet som i Eksempel 3.3.

Verdien $G_{c,d}$ 15 blir klassifisert til litelogen Stein og det kontrolleres om den ble klassifisert til riktig litelogi. Dette gjentas for alle verdiene $G_{c,d}$ til Sand og Stein. Resultatene blir presentert som vist i forvirringsmatrisen i Figur 4.1.

| | Sand | Stein |
|-------|------|-------|
| Sand | 6 | 0 |
| Stein | 0 | 6 |

Figur 4.1: Eksempel på forvirringsmatrise

I Figur 4.1 kan det ses at alle verdiene $G_{c,d}$ ble klassifisert riktig i dette eksempelet. Årsaken til det er at verdiene $G_{c,d}$ ligger i to klart atskilte verdiområder. Om verdien $G_{c,d}$ 15 hadde blitt klassifisert til Sand, ville antallet i den markerte cellen i Figur 4.1 blitt øket med én.

4.2 Klassifisering ved hjelp av én sensor

Vi undersøkte hvordan verdiområdene til litelogene for hver sensor lå fordelt, og det viste seg at de lå på en form som var tilnærmet lik en normalfordeling. Ut i fra denne observasjonen ble det valgt å benytte normalfordeling til å skille litelogene.

I de fire loggene som har blitt valgt ut som et representativt utvalg, har det blitt sett på hvor godt verdiene $G_{c,d}$ kan klassifiseres ved å kun se på én enkelt sensor. Vi tar utgangspunkt i alle sensorene som egner seg til å skille litelogene i hver enkelt logg, som forklart i Kapittel 2.2. Det blir sett på hvor godt hver enkelt sensor klarer å klassifisere litelogene.

Dette delkapittelet presenterer hvilken sensor som er best, og hvilken som er dårligst til å skille litelogene i hver logg. Sensorene i loggene har blitt undersøkt ved hjelp av leave-one-out metoden.

I loggene som nevnes i Kapittel 2 er verdier $G_{c,d}$ som er satt til -999,25, altså dummyverdier, sett bort i fra. Dette er årsaken til at antall målte verdier $G_{c,d}$ til en sensor kan være ulikt i forhold til de andre sensorene i loggen.

4.2.1 Logg 1

Forvirringsmatrisene i denne rapporten ligger på et fast format med et fast innhold. Til høyre i tabellen representerer feltet *Alle*, antallet verdier $G_{c,d}$ målt av én sensor i loggen. Klassifiseres litelogene med flere sensorer, representerer feltet antall målinger \vec{G}_d . Feltet *Riktig* beskriver antallet som er klassifisert til riktig litelogi. I *Feil*, finnes det antallet som blir klassifisert til feil litelogi. Feltene *Feilrate* og *Klassifisering* inneholder henholdsvis sannsynligheten for feil og riktig klassifisering.

Logg 1 er den første loggen som har blitt brukt som utgangspunkt for å undersøke sensorene. Sensoren som har best klassifiseringssannsynlighet er HALS Measurements / Medium Resistivity, og sannsynligheten for riktig klassifisering er 0,7529. Se Tabell 4.1 for forvirringsmatrisen.

Tabell 4.1: Forvirringsmatrise for den beste sensoren i Logg 1

| | Sandstone | Limestone | Claystone | Siltstone |
|-----------|-----------|-----------|-----------|-----------|
| Sandstone | 1125 | 178 | 952 | 66 |
| Limestone | 40 | 54 | 14 | 97 |
| Claystone | 0 | 59 | 8980 | 1936 |
| Siltstone | 0 | 0 | 17 | 78 |

Alle: 13596
Riktig: 10237
Feil: 3359
Feilrate: 0,24705796
Klassifisering: 0,75294204

Sensoren som klassifiserer litelogiene dårligst i Logg 1 er Thermal Neutron Porosity. Sannsynligheten for riktig klassifisering er 0,5916, og forvirringsmatrisen til sensoren er vist i Tabell 4.2.

Tabell 4.2: Forvirringsmatrise for den dårligste sensoren i Logg 1

| | Sandstone | Limestone | Claystone |
|-----------|-----------|-----------|-----------|
| Sandstone | 1041 | 186 | 626 |
| Limestone | 0 | 172 | 33 |
| Claystone | 10 | 46 | 92 |

Alle: 2206
Riktig: 1305
Feil: 901
Feilrate: 0,40843155
Klassifisering: 0,59156845

Dette betyr at sensorene i Logg 1 har en klassifiseringssannsynlighet som går fra 0,5916 til 0,7529. Forvirringsmatrisene og klassifiseringssannsynlighetene til alle sensorene i denne loggen finnes i Vedlegg A.

4.2.2 Logg 2

Den andre loggen det har blitt tatt utgangspunkt i er Logg 2. Denne loggen har kun to litelogier å skille blant, Sandstone og Claystone. Med dette som utgangspunkt vil det være naturlig å anta at klassifisering vil kunne være noe bedre enn for Logg 1 siden det er færre litelogier som kan ligge i samme verdiområde.

I Logg 2 klassifiserer HALS Measurements / Medium Resistivity litelogiene best, og sensoren hadde en klassifiseringssannsynlighet på 0,9973. Dette vil si at cirka 99,73 % av alle verdier $G_{c,d}$ blir klassifisert til riktig litelogi, se Tabell 4.3 for resultater.

Tabell 4.3: Forvirringsmatrise for den beste sensoren i Logg 2

| | Sandstone | Claystone |
|-----------|-----------|-----------|
| Sandstone | 30642 | 14 |
| Claystone | 86 | 6746 |

Alle: 37488
Riktig: 37388
Feil: 100
Feilrate: 0,00266752
Klassifisering: 0,99733248

Sensoren som klassifiserer litelogiene dårligst i Logg 2 er, Medium Shallow Resistivity. Denne klassifiserte litelogiene med en sannsynlighet på 0,3404, se Tabell 4.4 for forvirringsmatrisen til Medium Shallow Resistivity.

Tabell 4.4: Forvirringsmatrise for den dårligste sensoren i Logg 2

| | Sandstone | Claystone |
|-----------|-----------|-----------|
| Sandstone | 4872 | 10165 |
| Claystone | 24 | 386 |

Alle: 15447
Riktig: 5258
Feil: 10189
Feilrate: 0,65961028
Klassifisering: 0,34038972

Dette fører til at i Logg 2 klassifiseres litelogiene med en sannsynlighet i området 0,3404 til 0,9973 for de sju sensorene. Ut i fra dette kan det ses at den klassifiseres noe bedre enn Logg 1, om det velges å ta utgangspunkt i den beste sensoren. Blir sensoren med lavest klassifiseringssannsynlighet valgt, klassifiseres litelogiene langt dårligere enn i Logg 1. Se Vedlegg B for resultatene til de andre sensorene i Logg 2.

4.2.3 Logg 3

Logg 3 var den neste loggen som ble brukt som utgangspunkt. Denne inneholder de tre litelogiene Sandstone, Limestone og Claystone, og har fem forskjellige sensorer som egner seg til å skille litelogiene.

HALS Measurements / Medium Resistivity er den sensoren som skiller litelogiene best med en klassifiseringssannsynlighet på 0,7918, se Tabell 4.5.

Tabell 4.5: Forvirringsmatrise for den beste sensoren i Logg 3

| | Sandstone | Limestone | Claystone |
|-----------|-----------|-----------|-----------|
| Sandstone | 121 | 768 | 2115 |
| Limestone | 16 | 10 | 363 |
| Claystone | 0 | 53 | 12476 |

Alle: 15922
Riktig: 12607
Feil: 3315
Feilrate: 0,20820249
Klassifisering: 0,79179751

Sensoren som klassifiserer litelogiene dårligst i Logg 3 er High Resolution Bulk Density, og den klassifiserer med en sannsynlighet 0,3772. Se Tabell 4.6 for resultatene til denne sensoren.

Tabell 4.6: Forvirringsmatrise for den dårligste sensoren i Logg 3

| | Sandstone | Limestone | Claystone |
|-----------|-----------|-----------|-----------|
| Sandstone | 921 | 580 | 1370 |
| Limestone | 25 | 124 | 166 |
| Claystone | 3 | 21 | 266 |

Alle: 3476
Riktig: 1311
Feil: 2165
Feilrate: 0,62284235
Klassifisering: 0,37715765

Ved å benytte én sensor til å skille litelogiene, vil sannsynligheten for å klassifisere litelogiene gå fra 0,3772 til 0,7918. Igjen er det stor forskjell på hvor godt litelogiene blir klassifisert avhengig av hvilken sensor det blir tatt utgangspunkt i. Forvirringsmatrisene og klassifiseringssannsynlighetene til alle sensorene i Logg 3, finnes i Vedlegg C.

4.2.4 Logg 4

I den siste loggen som har blitt undersøkt er det funnet fire forskjellige litologier, og fem sensorer som kan brukes til å skille litologiene.

Thermal Neutron Porosity er sensoren som klassifiserer litologiene beste i denne loggen, og har en klassifiseringssannsynlighet på 0,9894. Se forvirringsmatrisen i Tabell 4.7.

Tabell 4.7: Forvirringsmatrise for den beste sensoren i Logg 4

| | Sandstone | Limestone | Shale | Anhydrite |
|-----------|-----------|-----------|-------|-----------|
| Sandstone | 214 | 4 | 0 | 0 |
| Limestone | 1 | 450 | 0 | 1 |
| Shale | 0 | 0 | 0 | 0 |
| Anhydrite | 0 | 3 | 0 | 174 |

Alle: 847
Riktig: 838
Feil: 9
Feilrate: 0,01062574
Klassifisering: 0,98937426

I Logg 4 er den sensoren som klassifiserer litologiene med høyest feilrate, High Resolution Gamma Ray, og sannsynligheten er 0,8967. Dette vil si at 89,67 % av alle verdiene $G_{c,d}$ blir klassifisert til riktig litologi, se Tabell 4.8 for resultater.

Tabell 4.8: Forvirringsmatrise for den dårligste sensoren i Logg 4

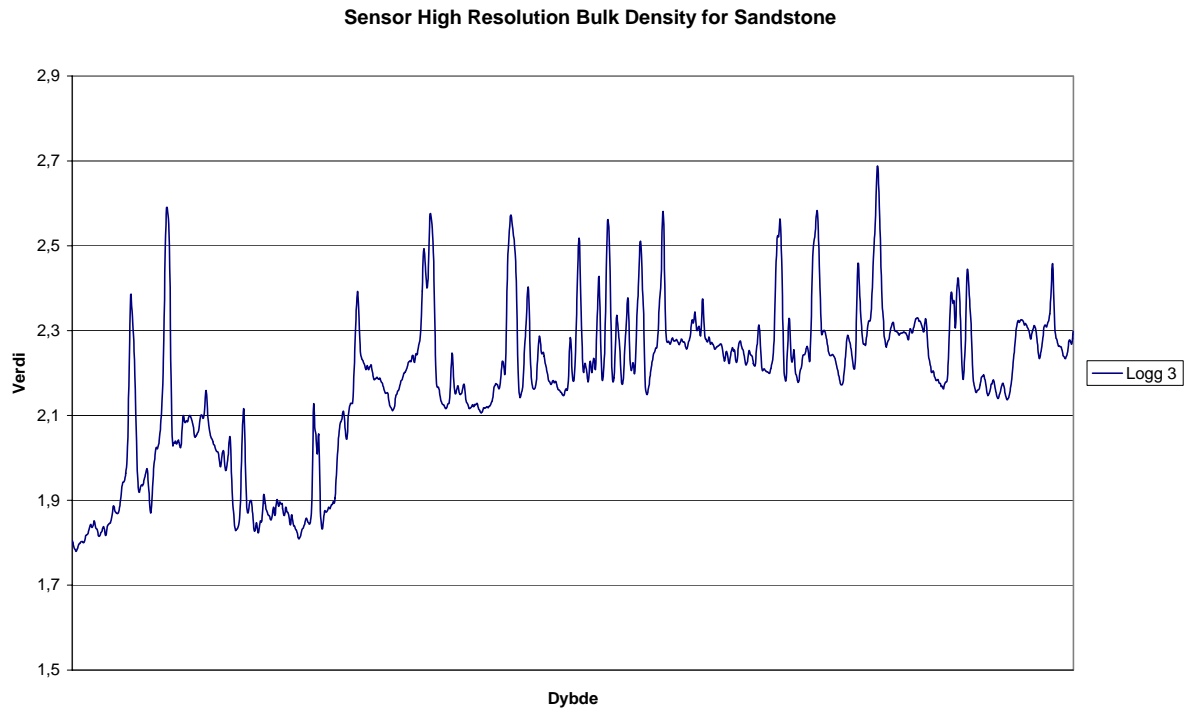
| | Sandstone | Limestone | Shale | Anhydrite |
|-----------|-----------|-----------|-------|-----------|
| Sandstone | 213 | 5 | 0 | 0 |
| Limestone | 21 | 284 | 0 | 147 |
| Shale | 0 | 0 | 1877 | 0 |
| Anhydrite | 1 | 111 | 0 | 101 |

Alle: 2760
Riktig: 2475
Feil: 285
Feilrate: 0,10326087
Klassifisering: 0,89673913

Dette betyr at litologiene blir klassifisert med en sannsynlighet fra 0,8967 til 0,9894. I Vedlegg D finnes forvirringsmatrisene og klassifiseringssannsynlighetene til de resterende tre sensorene i denne loggen.

4.2.5 Verdiområdet ved boreddybde

Vi gjorde også en undersøkelse om verdiområdet, som kjennetenger litologiene, forandret seg avhengig av boreddybde. Undersøkelsen av de fire loggene viste at verdiene $G_{c,d}$ lå i samme verdiområde uavhengig av dybden, se Figur 4.2 for et eksempel på dette.

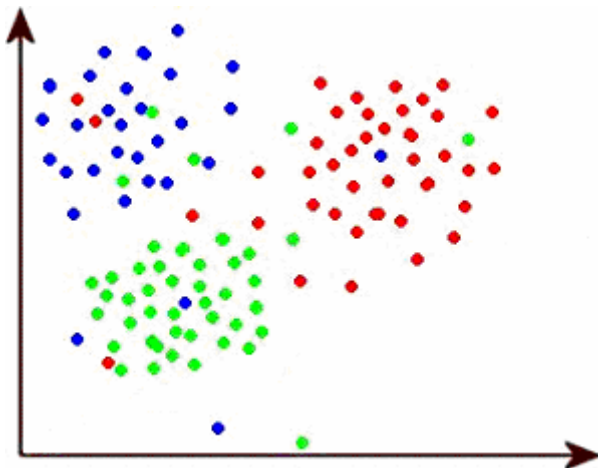


Figur 4.2: Verdiområdet ved dybde

Figuren viser at det er en viss forandring i verdiområdet under boringen, men dette skyldes støy og er ikke knyttet til boreddybden.

4.3 Klassifisering ved hjelp av Naive Bayes Classifier

I forrige delkapittel ble det sett på hvor godt man klarer å klassifisere litologiene ved å se på kun én sensor. Det neste naturlige steget er å se om en bedre klassifisering kan oppnås ved å benytte seg av alle sensorene til å skille litologiene. Ved å benytte flere sensorer, vil dette tilsvare å klassifisere over flere dimensjoner. Trolig vil det være lettere å skille litologiene jo flere sensorer som blir benyttet, og Figur 4.3 illustrerer dette.



Figur 4.3: Eksempel på målinger over flere dimensjoner

Av figuren kan det ses at noen målinger \vec{G}_d vil kunne bli feilklassifisert ved at de ligger i verdiområdet til en annen litelogi. Ved å benytte seg av flere målinger, vil muligens noen av disse feilene kunne bli rettet opp slik at de blir klassifisert til riktig litelogi.

For å undersøke hvor godt litelogiene kan skilles ved å benytte flere sensorer valgte vi å bruke Naive Bayes Classifier. Årsaken til at nettopp denne metoden ble valgt skyldes metodens egnede egenskaper. Naive Bayes Classifier tolererer støy i datasettet, samtidig som den gir en gradering av hvor sikker klassifiseringen er. En annen fordel med Naive Bayes Classifier er at systemet trenes opp inkrementelt. Det vil si at metoden ikke krever at alle treningsdataene blir lest inn på en gang. I tillegg er minne og prosesseringsforbruk konstant.

Vi undersøkte loggene ved hjelp av leave-one-out metoden, se Kapittel 4.1 for forklaring. For å gi et bilde av hvor godt Naive Bayes Classifier klassifiserer litelogiene, vil resultatene bli illustrert ved hjelp av en forvirringsmatrise i hver logg.

4.3.1 Logg 1

Logg 1 har seks sensorer som egnet seg til å klassifisere litelogiene. Ved å benytte de sensorene i metoden Naive Bayes Classifier, ble resultatet som Tabell 4.9 viser.

Tabell 4.9: Forvirringsmatrise for Naive Bayes Classifier i Logg 1

| | Sandstone | Limestone | Claystone | Siltstone | |
|-----------|-----------|-----------|-----------|-----------|-----------------------------------|
| Sandstone | 1169 | 111 | 956 | 88 | Alle: 13638 |
| Limestone | 41 | 80 | 84 | 0 | Riktig: 11880 |
| Claystone | 10 | 44 | 10537 | 423 | Feil: 1758 |
| Siltstone | 0 | 0 | 1 | 94 | Feilrate: 0,12890453 |
| | | | | | Klassifisering: 0,87109547 |

Av tabellen kan det ses at litelogiene blir klassifisert med en sannsynlighet tilnærmet lik 0,8711, og det vil si at Naive Bayes Classifier klarer å klassifisere 87,11 % av alle målinger \vec{G}_d til riktig litelogi.

4.3.2 Logg 2

Den andre loggen som det har blitt tatt utgangspunkt i, har sju forskjellige sensorer som eger seg til bruk i klassifiseringsmetoden. Ved å benytte Naive Bayes Classifier med de sensorene, ble resultatene som forvirringsmatrisen i Tabell 4.10 viser.

Tabell 4.10: Forvirringsmatrise for Naive Bayes Classifier i Logg 2

| | Sandstone | Claystone | Limestone | |
|-----------|-----------|-----------|-----------|------------------------------------|
| Sandstone | 30993 | 1612 | 375 | Alle: 39826 |
| Claystone | 82 | 6764 | 0 | Riktig: 37757 |
| | | | | Feil: 2069 |
| | | | | Feilrate: 0,051950987 |
| | | | | Klassifisering: 0,948049013 |

Tabell 4.10 viser at litelogiene blir klassifisert med en sannsynlighet på 0,9480. Det vil med andre ord si at 94,80 % av alle målingene \vec{G}_d blir klassifisert til riktig litelogi.

En ting som er verdt å merke seg, og som kan virke merkelig ved første øyekast, er at flere målinger \vec{G}_d blir klassifisert til litelogien Limestone. Denne litelogien finnes ikke i loggen,

men det er likevel en naturlig forklaring på at noen målinger \vec{G}_d kan bli klassifisert til denne litelogien.

Alle litelogier som er representert i en eller flere av loggene, er det mulig for Naive Bayes Classifier å klassifisere målingene \vec{G}_d til. Dersom en litelogi ikke er representert i loggen, og dermed heller ikke kan trenes opp på vanlig måte, vil denne bli trent opp etter visse fastlagte kriterier. Litelogier uten treningsgrunnlag får middelverdien satt som snittet av middelverdiene til de litelogiene som er blitt trent opp. Standardavviket til litelogiene som mangler treningsgrunnlag blir satt til 100 ganger høyere enn det største standardavviket til litelogiene som er representert i loggen.

Det er dermed mulig at målinger \vec{G}_d blir klassifisert til litelogier som ikke finnes i loggen. Grunnen til at det nettopp står Limestone og ikke en av de andre litelogiene som heller ikke er representert i loggen, er at alle litelogier som ikke finnes i loggen alltid vil ha samme sannsynlighet for å slå til. Dersom Naive Bayes Classifier anser det som mest sannsynlig at det er en av disse litelogiene. Velges den første og beste litelogien til å representere at målingen \vec{G}_d har blitt klassifisert til en litelogi som ikke finnes i loggen. I dette tilfellet er denne litelogien Limestone. Dette er en utvidelse i forhold til klassifiseringen gjort med én sensor.

4.3.3 Logg 3

Den neste loggen som vi undersøkte inneholdt fem forskjellige sensorer som kan benyttes i Naive Bayes Classifier. Når vi benyttet disse sensorene i klassifiseringsmetoden fikk vi resultatene som Tabell 4.11 viser.

Tabell 4.11: Forvirringsmatrise for Naive Bayes Classifier i Logg 3

| | Sandstone | Limestone | Claystone | Shale |
|-----------|-----------|-----------|-----------|-------|
| Sandstone | 831 | 338 | 1831 | 4 |
| Limestone | 25 | 196 | 229 | 0 |
| Claystone | 28 | 289 | 12227 | 0 |

| | |
|------------------------|------------|
| Alle: | 15998 |
| Riktig: | 13254 |
| Feil: | 2744 |
| Feilrate: | 0,17152144 |
| Klassifisering: | 0,82847856 |

Tabell 4.11 viser at sannsynligheten til å klassifisere litelogiene ved hjelp av Naive Bayes Classifier er på 0,8285. Det vil si at klassifiseringsmetoden klarer å klassifisere 82,85 % av alle målingene \vec{G}_d til riktig litelogi.

4.3.4 Logg 4

Den siste loggen det har blitt sett nærmere på, har fem sensorer som Naive Bayes Classifier kan benytte til å klassifisere litelociene. Ved å benytte disse sensorene ble resultatene som Tabell 4.12 viser.

Tabell 4.12: Forvirringsmatrise for Naive Bayes Classifier i Logg 4

| | Sandstone | Limestone | Shale | Anhydrite | Claystone |
|-----------|-----------|-----------|-------|-----------|-----------|
| Sandstone | 214 | 0 | 0 | 0 | 3 |
| Limestone | 0 | 446 | 0 | 3 | 2 |
| Shale | 0 | 0 | 1875 | 0 | 0 |
| Anhydrite | 0 | 0 | 0 | 211 | 4 |

Alle: 2758
Riktig: 2746
Feil: 12
Feilrate: 0,00435098
Klassifisering: 0,99564902

Tabellen viser at Naive Bayes Classifier gir en sannsynlighet for riktig klassifisering som er tilnærmet lik 0,9956. Det vil si at cirka 99,56 % av alle målingene \vec{g}_d klassifiseres til riktig liteloci.

4.4 Klassifisering ved hjelp av Hidden Markov Model

Etter å ha sett hvor godt litelociene blir klassifiserte ved hjelp av Naive Bayes Classifier, er neste skritt å se om det er en korrelasjon mellom målingene \vec{g}_d . Det vil si om målingene \vec{g}_d kan klassifiseres bedre ved å se på før og etterliggende målinger \vec{g}_d . Metoden som vi har valgt å benytte for å undersøke dette er Hidden Markov Model. Denne bygger på Naive Bayes Classifier og har derfor de samme fordelene, men tar i tillegg hensyn til før og etterliggende målinger \vec{g}_d i tid / rom.

I disse eksperimentene har vi en egen overgangsmatrise per logg som er spesiallaget for hver logg. Overgangsmatrisene blir laget ved at man teller opp antall skift, fra en liteloci til enten seg selv eller en annen, i loggen. Når dette er gjort, regnes det ut sannsynligheten for at et skifte vil inntreffe. Se Figur 3.2 for et eksempel på hvordan en slik overgangsmatrise blir satt opp.

Også i dette delkapittelet undersøker vi hvor godt Hidden Markov Model klassifiserer målingene \vec{g}_d logg for logg. Resultatene blir fremstilt i forvirringsmatriser for alle loggene.

4.4.1 Logg 1

Ved å benytte Hidden Markov Model som klassifiseringsmetode, ble litelociene klassifisert som Tabell 4.13 viser.

Tabell 4.13: Forvirringsmatrise for Hidden Markov Model på Logg 1

| | Sandstone | Limestone | Claystone | Siltstone |
|-----------|-----------|-----------|-----------|-----------|
| Sandstone | 1173 | 106 | 980 | 65 |
| Limestone | 41 | 79 | 85 | 0 |
| Claystone | 11 | 40 | 10659 | 304 |
| Siltstone | 0 | 0 | 0 | 95 |

Alle: 13638
Riktig: 12006
Feil: 1632
Feilrate: 0,11966564
Klassifisering: 0,88033436

Tabell 4.13 viser at Hidden Markov Model klassifiserer litelociene i Logg 1 med en sannsynlighet som er tilnærmet lik 0,8803. Det vil si at cirka 88,03 % av alle målingene \vec{g}_d blir klassifisert til riktig liteloci.

4.4.2 Logg 2

Når Hidden Markov Model ble benyttet som klassifiseringsmetode i Logg 2, ble resultatene som forvirringsmatrisen i Tabell 4.14 illustrerer.

Tabell 4.14: Forvirringsmatrise for Hidden Markov Model på Logg 2

| | Sandstone | Claystone | Limestone |
|-----------|-----------|-----------|-----------|
| Sandstone | 31070 | 1548 | 362 |
| Claystone | 75 | 6771 | 0 |

Alle: 39826
Riktig: 37841
Feil: 1985
Feilrate: 0,04984181
Klassifisering: 0,95015819

Ut i fra tabellen leses det at Hidden Markov Model klassifiserer litelogiene med en sannsynlighet på cirka 0,9502. Omregnet til prosent tilsvarer dette at 95,02 % av alle målingene \vec{g}_a i Logg 2 blir klassifisert til riktig litelogi.

4.4.3 Logg 3

Hidden Markov Model produserte resultatene som Tabell 4.15 viser når litelogiene i Logg 3 ble klassifisert.

Tabell 4.15: Forvirringsmatrise for Hidden Markov Model på Logg 3

| | Sandstone | Limestone | Claystone | Shale |
|-----------|-----------|-----------|-----------|-------|
| Sandstone | 834 | 337 | 1830 | 3 |
| Limestone | 16 | 193 | 241 | 0 |
| Claystone | 14 | 260 | 12270 | 0 |

Alle: 15998
Riktig: 13297
Feil: 2701
Feilrate: 0,1688336
Klassifisering: 0,8311664

Når Hidden Markov Model blir brukt som klassifiseringsmetode blir litelogiene klassifisert med en sannsynlighet som er tilnærmet lik 0,8312. Det vil si at hvis Hidden Markov Model blir benyttet på Logg 3, blir cirka 83,12 % av alle målingene \vec{g}_a klassifisert til riktig litelogi.

4.4.4 Logg 4

Den siste loggen som Hidden Markov Model ble testet ut på, produserte resultatene som forvirringsmatrisen i Tabell 4.16 viser.

Tabell 4.16: Forvirringsmatrise for Hidden Markov Model på Logg 4

| | Sandstone | Limestone | Shale | Anhydrite | Claystone |
|-----------|-----------|-----------|-------|-----------|-----------|
| Sandstone | 214 | 0 | 0 | 0 | 3 |
| Limestone | 0 | 444 | 0 | 5 | 2 |
| Shale | 0 | 0 | 1875 | 0 | 0 |
| Anhydrite | 0 | 0 | 0 | 211 | 4 |

| | |
|------------------------|------------|
| Alle: | 2758 |
| Riktig: | 2744 |
| Feil: | 14 |
| Feilrate: | 0,00507614 |
| Klassifisering: | 0,99492386 |

Tabellen viser at ved å benytte Hidden Markov Model oppnås en klassifiseringssannsynlighet som er tilnærmet lik 0,9949. Det vil i praksis si at Hidden Markov Model klassifiserer cirka 99,49 % av alle målingene \vec{g}_d til riktig litologi.

Om Hidden Markov Model benyttes til å klassifisere litologiene på en boreplattform, vil ikke en nøyaktig overgangsmatrise kunne lages. Årsaken til dette er at man ikke vet på forhånd hvordan grunnen ser ut der det skal bores, med tanke på skift mellom litologiene. En generell snittovergangsmatrise som kan benyttes uansett hvor det bores, bør derfor lages.

For å undersøke om dette påvirker hvor godt Hidden Markov Model klarer å klassifisere litologiene, lagde vi en ny overgangsmatrise som var et snitt av de spesiallagde overgangsmatrisene til de fire loggene. Deretter forsøkte vi å klassifisere loggene på nytt med Hidden Markov Model og den nye snittovergangsmatrisen.

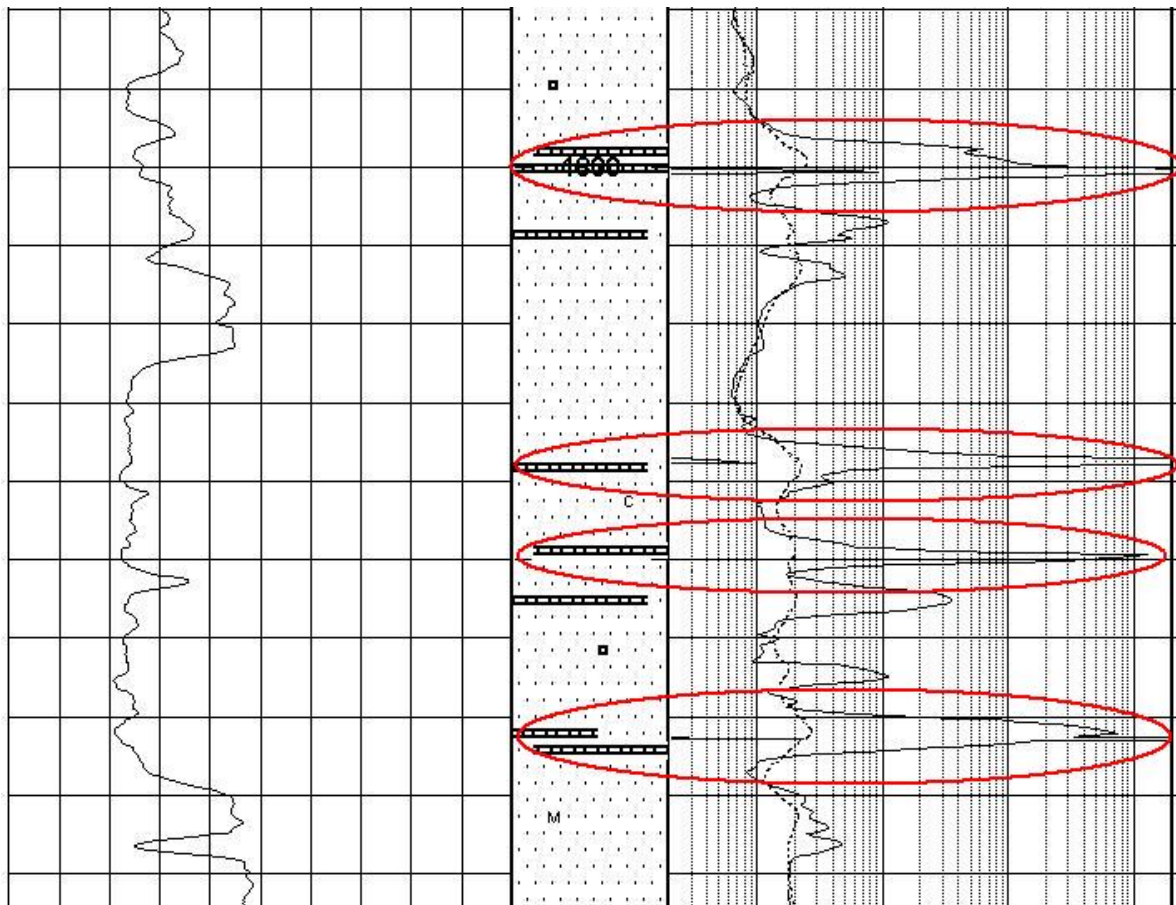
Resultatene av dette forsøket, var at i Logg 1 ble litologiene klassifiserte med en sannsynlighet 0,8806. I Logg 2 ble litologiene klassifisert med en sannsynlighet på 0,9488. Når Hidden Markov Model ble benyttet med den nye snittovergangsmatrisen på Logg 3, fikk vi klassifiseringssannsynligheten 0,8312. Resultatene fra klassifiseringen av Logg 4 når den nye snittovergangsmatrisen ble benyttet, var at litologiene ble klassifisert med en sannsynlighet tilnærmet lik 0,9949. Forvirringsmatrisene og nøyaktig klassifiseringssannsynlighet for alle loggene med den nye snittovergangsmatrisen, finnes i loggenes respektive vedlegg. Ut i fra resultatene blir ikke klassifiseringssannsynlighetene påvirket i større grad. I videre undersøkelser hvor Hidden Markov Model blir benyttet er de spesiallagde overgangsmatrisene brukt, om ikke noe annet er nevnt.

4.5 Utbedringer av klassifiseringen

I dette delkapittelet blir det sett på forskjellige optimaliseringer av klassifiseringsmetodene og treningsdataene. Hensikten er å bedre klassifiseringssannsynligheten.

4.5.1 Fjerne uteliggere

Det første som ble undersøkt, når vi så på hva som kunne forbedre klassifiseringen, var å avgjøre i hvor stor grad sedimentene påvirker graden av riktig klassifisering. Ved å se på litelogiprofilene til de klassifiserte loggene, kunne det tyde på at sedimentene påvirket verdiene $G_{c,d}$ som blir målt av sensorene.



Figur 4.4: Utsnitt av en logg som viser innslag av sedimenter

Figur 4.4 er et utsnitt av en litelogiprosent, hvor grafen til venstre viser gammastrålingen. I midten av figuren er litologiene illustrert med innslag av sedimenter, og grafene til høyre i figuren anviser hvordan resistiviteten forandrer seg. Grafene viser at resistiviteten blir påvirket av sedimentene. Det skjer store hopp i resistivitetsverdiene $G_{c,d}$ når det støtes på sedimenter, disse hoppene er avmerket med en rød sirkel i figuren.

En undersøkelse for å se i hvor stor grad sedimentene påvirket klassifiseringen til Naive Bayes Classifier og Hidden Markov Model ble utført. Ved at det skjer store endringer i verdiene $G_{c,d}$ målt av sensoren, vil dette føre til at treningen av litologiene blir påvirket. Middelerdien vil bli satt til en annen verdi enn hva litologien egentlig burde hatt. Ikke minst

vil store endringer i verdiene $G_{c,d}$ føre til at standardavviket vil bli mye større enn hva det ville ha blitt dersom sedimentene ikke var til stede. I verste fall kan disse store standardavvikene føre til feilklassifiseringer.

For å undersøke i hvor stor grad sedimentene påvirket klassifiseringssannsynligheten, ble det fjernet 1 % av de høyeste verdiene $G_{c,d}$ i alle sensorene for alle litelogiene. De høye verdiene $G_{c,d}$ blir også kalt uteliggere. Resultatene fra Logg 1 og 2 blir vist, mens de øvrige resultatene finnes i Vedlegg C og D.

I Logg 1 ble 1 % av uteliggerne fjernet og deretter klassifisert på nytt. Når klassifiseringen ble utført med Naive Bayes Classifier ved hjelp av det nye treningsgrunnlaget, ble resultatene som forvirringsmatrisen i Tabell 4.17 viser.

Tabell 4.17: Logg 1 klassifisert med Naive Bayes Classifier uten 1 % av uteliggerne

| | Sandstone | Limestone | Claystone | Siltstone | | |
|-----------|-----------|-----------|-----------|-----------|------------------------|------------|
| Sandstone | 1169 | 158 | 922 | 74 | Alle: | 13637 |
| Limestone | 39 | 114 | 52 | 0 | Riktig: | 11857 |
| Claystone | 46 | 94 | 10479 | 395 | Feil: | 1780 |
| Siltstone | 0 | 0 | 0 | 95 | Feilrate: | 0,13052724 |
| | | | | | Klassifisering: | 0,86947276 |

Ved å sammenlikne tabellen over med Tabell 4.9, kan det ses at den totale klassifiseringssannsynligheten blir lavere enn hva den var originalt. Den ønskede effekten blir ikke oppnådd ved å fjerne 1 % av uteliggerne når Naive Bayes Classifier blir benyttet som klassifiseringsmetode. Det ble derfor sett på om effekten ble den samme når Hidden Markov Model ble benyttet. Resultatet blir vist i Tabell 4.18.

Tabell 4.18: Logg 1 klassifisert med Hidden Markov Model uten 1 % av uteliggerne

| | Sandstone | Limestone | Claystone | Siltstone | | |
|-----------|-----------|-----------|-----------|-----------|------------------------|------------|
| Sandstone | 1172 | 145 | 926 | 80 | Alle: | 13637 |
| Limestone | 41 | 109 | 55 | 0 | Riktig: | 11951 |
| Claystone | 40 | 85 | 10575 | 314 | Feil: | 1686 |
| Siltstone | 0 | 0 | 0 | 95 | Feilrate: | 0,12363423 |
| | | | | | Klassifisering: | 0,87636577 |

Ikke uventet kan det ses av Tabell 4.18 at også klassifiseringssannsynligheten til Hidden Markov Model blir lavere ved å fjerne 1 % av uteliggerne. Dette ser man ved å sammenlikne resultatet med Tabell 4.13.

Videre undersøkte vi om det var slik at klassifiseringssannsynligheten ble lavere i alle loggene ved å fjerne 1 % av uteliggerne. Vi undersøkte derfor om resultatene ble de samme i Logg 2. Naive Bayes Classifier klassifiserte loggen med det nye treningsgrunnlaget og fikk følgende resultater, se Tabell 4.19

Tabell 4.19: Logg 2 klassifisert med Naive Bayes Classifier uten 1 % av uteliggerne

| | Sandstone | Claystone | Limestone |
|-----------|-----------|-----------|-----------|
| Sandstone | 32235 | 32 | 710 |
| Claystone | 159 | 6687 | 0 |

Alle: 39823
Riktig: 38922
Feil: 901
Feilrate: 0,02262512
Klassifisering: 0,97737488

Sammenlikningen av Tabell 4.19 og Tabell 4.10 viser at fjerningen av uteliggerne fører til en betydelig bedring av klassifiseringssannsynligheten. Hidden Markov Model ble deretter benyttet som klassifiseringsmetode med det nye treningsgrunnlaget. Resultatet av dette forsøket blir vist i forvirringsmatrisen i Tabell 4.20.

Tabell 4.20: Logg 2 klassifisert med Hidden Markov Model uten 1 % av uteliggerne

| | Sandstone | Claystone | Limestone |
|-----------|-----------|-----------|-----------|
| Sandstone | 32221 | 23 | 733 |
| Claystone | 143 | 6703 | 0 |

Alle: 39823
Riktig: 38924
Feil: 899
Feilrate: 0,02257489
Klassifisering: 0,97742511

Av tabellen fremkommer det at også denne klassifiseringssannsynligheten blir mye bedre enn hva den var originalt. Dette kan ses ved å sammenlikne Tabell 4.20 og Tabell 4.14.

Den samme undersøkelsen ble gjort på alle de fire loggene. Resultatene fra Logg 3 og 4 var at i Logg 3 ble klassifiseringssannsynlighet dårligere ved å fjerne 1 % av uteliggerne, mens i Logg 4 ble den bedre. Ut i fra disse resultatene kan det ikke anbefales å fjerne 1 % av de høyeste verdiene $G_{c,d}$ til hver sensor i hver litologi.

Vi valgte derfor å gjøre en undersøkelse om klassifiseringssannsynligheten ble bedre ved å fjerne 1 % av de høyeste verdiene $G_{c,d}$ per sensor for hver litologi. Resultatene til de samme loggene som over, blir vist for å illustrere resultatene fra henholdsvis Naive Bayes Classifier og Hidden Markov Model.

Logg 1 var den første loggen hvor det ble forsøkt å klassifisere litologiene uten 1 % av uteliggerne. Ved å benytte Naive Bayes Classifier som klassifiseringsmetode, ble resultatene som Tabell 4.21 viser.

Tabell 4.21: Logg 1 klassifisert med Naive Bayes Classifier uten 1 % av uteliggerne

| | Sandstone | Limestone | Claystone | Siltstone |
|-----------|-----------|-----------|-----------|-----------|
| Sandstone | 1175 | 141 | 926 | 81 |
| Limestone | 41 | 102 | 62 | 0 |
| Claystone | 12 | 59 | 10578 | 365 |
| Siltstone | 0 | 0 | 1 | 94 |

Alle: 13637
Riktig: 11949
Feil: 1688
Feilrate: 0,12378089
Klassifisering: 0,87621911

Ved å sammenlikne resultatene fra Tabell 4.21 med resultatene fra det originale datasettet, ser man at en høyere klassifiseringssannsynlighet blir oppnådd ved å fjerne 1 % av uteliggerne. På denne loggen ble det for mye å fjerne 1 % av uteliggerne, ved å fjerne kun 1 % ble den ønskede effekten oppnådd.

Dersom Hidden Markov Model blir benyttet som klassifiseringsmetode med det nye treningsgrunnlaget, blir resultatene som forvirringsmatrisen i Tabell 4.22 viser.

Tabell 4.22: Logg 1 klassifisert med Hidden Markov Model uten 1 % av uteliggerne

| | Sandstone | Limestone | Claystone | Siltstone |
|-----------|-----------|-----------|-----------|-----------|
| Sandstone | 1178 | 130 | 935 | 80 |
| Limestone | 41 | 102 | 62 | 0 |
| Claystone | 12 | 56 | 10675 | 271 |
| Siltstone | 0 | 0 | 0 | 95 |

Alle: 13637
Riktig: 12050
Feil: 1587
Feilrate: 0,11637457
Klassifisering: 0,88362543

Her blir også klassifiseringssannsynligheten bedre enn det som ble oppnådd med å klassifisere det originale datasettet.

Eksperimentet som ble utført på Logg 1 ble også utført på Logg 2, og Naive Bayes Classifier produserte da resultatene som blir vist i Tabell 4.23.

Tabell 4.23: Logg 2 klassifisert med Naive Bayes Classifier uten 1 % av uteliggerne

| | Sandstone | Claystone | Limestone |
|-----------|-----------|-----------|-----------|
| Sandstone | 31387 | 942 | 650 |
| Claystone | 100 | 6746 | 0 |

Alle: 39825
Riktig: 38133
Feil: 1692
Feilrate: 0,04248588
Klassifisering: 0,95751412

Tabell 4.23 viser at den totale klassifiseringssannsynligheten blir bedre enn den originale, men blir derimot ikke så bra som resultatene som oppnås ved å fjerne 1 % av uteliggerne.

Det ble også sett på resultatene som ble oppnådd når Hidden Markov Model ble benyttet som klassifiseringsmetode med det nye treningsgrunnlaget. Resultatene blir illustrert i Tabell 4.24 viser.

Tabell 4.24: Logg 2 klassifisert med Hidden Markov Model uten 1 % av uteliggerne

| | Sandstone | Claystone | Limestone |
|-----------|-----------|-----------|-----------|
| Sandstone | 31405 | 929 | 645 |
| Claystone | 84 | 6762 | 0 |

Alle: 39825
Riktig: 38167
Feil: 1658
Feilrate: 0,04163214
Klassifisering: 0,95836786

Forvirringsmatrisen i Tabell 4.24 viser at det blir oppnådd en høyere klassifiseringssannsynlighet enn for det originale datasettet. Ved å fjerne 1 % av uteliggerne, blir derimot ikke litelogene klassifisert like bra som om 1 % hadde blitt fjernet.

Eksperimentet med å fjerne 1 % av uteliggerne ble også utført på de to siste loggene. Resultatet i Logg 3 ble at klassifiseringssannsynlighet ble høyere enn det som ble oppnådd med det originale datasettet. I Logg 4 ble sannsynligheten for riktig klassifisering lik som i det originale datasettet. Forvirringsmatrisene og klassifiseringssannsynligheten ved å fjerne én promille av uteliggerne for Logg 3 og 4 finnes i henholdsvis Vedlegg C og D.

4.5.2 Fjerne sensorer

En annen mulighet til å forbedre klassifiseringen er å se på om alle sensorene gir et positivt utslag. Det vil si om alle sensorene bidrar til at målingene \vec{c}_i blir klassifisert til riktig litelogi. Ved å se på resultatene i Kapittel 4.2, ser man at enkelte sensorer er dårlige til å skille litelogiene. De vil muligens bidra til at litelogiene blir feilklassifisert, dersom klassifiseringsmetoden står og ”vipper” mellom to litelogier.

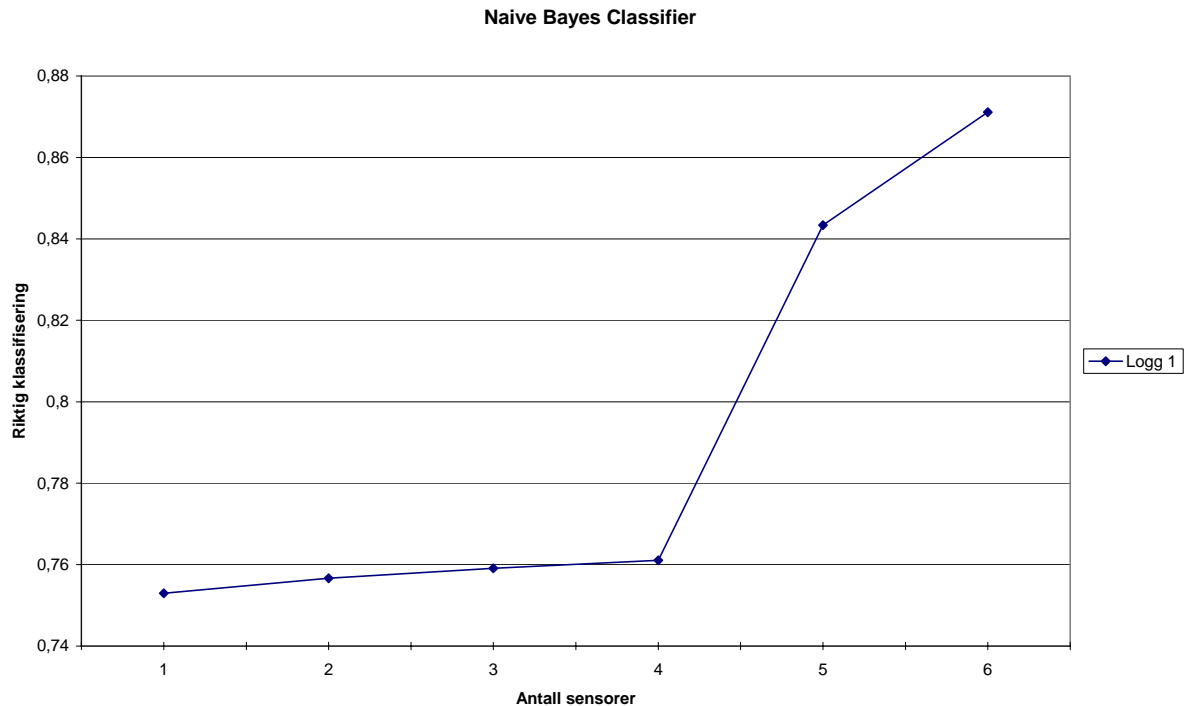
Vi utførte et eksperiment for å undersøke om en bedre klassifikasjonssannsynlighet kunne oppnås ved å fjerne enkelte sensorer. For å undersøke dette så vi på sensorene i alle loggene for å finne ut hvor godt sensorene egner seg til å skille litelogiene i snitt, og Tabell 4.25 viser dette.

Tabell 4.25: Liste over hvilke sensorer som skiller litelogiene best

| | |
|--|------------|
| Delta-T / Sonic | 0,98659420 |
| Shallow Resistivity | 0,89095177 |
| Hals Measurements / Medium Resistivity | 0,88243830 |
| Deep Resistivity | 0,77999731 |
| Neutron Porosity | 0,69511901 |
| Bulk Density | 0,65642832 |
| Gamma Ray | 0,64598458 |
| Medium Shallow Resistivity | 0,54265735 |

Vi benyttet klassifiseringsmetodene først med den sensoren som egner seg best til å skille litelogiene, Delta-T / Sonic i dette tilfellet, og deretter med de to beste. Slik fortsatte vi til alle sensorene hadde blitt benyttet i klassifiseringsmetoden. Ved å gjøre dette ser man om klassifiseringssannsynligheten blir bedre eller dårligere om en sensor blir tatt med i metoden. Om det viser seg at klassifiseringen blir dårligere i en logg, vil det tyde på at en høyere klassifiseringssannsynlighet kan bli oppnådd ved å utelate sensoren når litelogiene skal skilles i alle loggene.

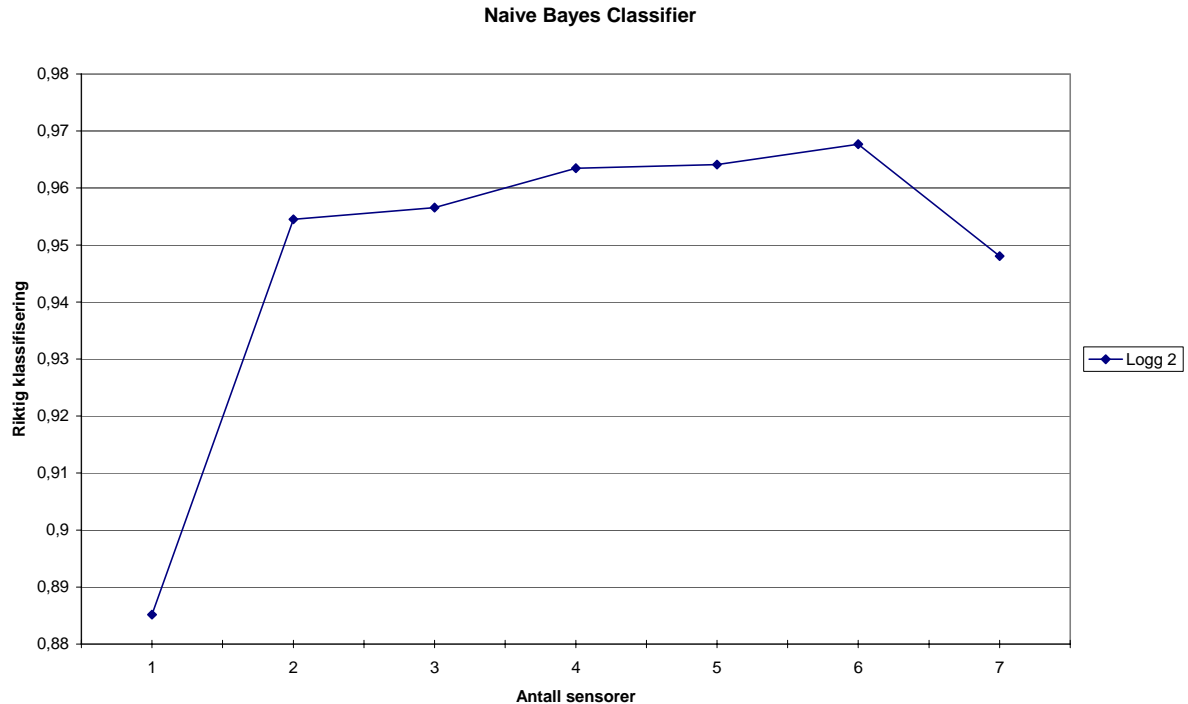
Når vi undersøkte Logg 1 fikk vi resultatene som grafen i Figur 4.5 representerer ved å benytte Naive Bayes Classifier.



Figur 4.5: Antall sensorer man bør benytte på Logg 1.

Figuren viser at klassifiseringssannsynligheten øker i takt med antall sensorer som blir benyttet. Dersom kun den sensoren som gir høyest klassifiseringssannsynlighet blir benyttet, blir litelogiene klassifisert med en sannsynlighet tilnærmet lik 0,7529. Den totale klassifiseringssannsynligheten der alle sensorene blir benyttet, er på cirka 0,8711. Ved å benytte Hidden Markov Model som klassifiseringsmetode, fås det en graf som er tilnærmet lik grafen som er illustrert i Figur 4.5. Denne grafen og forvirringsmatrisene kan studeres i Vedlegg A.

I Logg 2 blir resultatene noe annerledes, Figur 4.6 representerer resultatene for denne loggen når Naive Bayes Classifier blir benyttet som klassifiseringsmetode.



Figur 4.6: Antall sensorer man bør benytte på Logg 2.

Grafen i Figur 4.6 viser at klassifiseringssannsynlighet blir høyest om sensoren Medium Shallow Resistivity ikke blir benyttet. Ved å utelate denne sensoren, klassifiseres litelogiene med en sannsynlighet tilnærmet lik 0,9677. Inkluderes derimot denne sensoren, blir en sannsynlighet for riktig klassifisering av litelogiene på cirka 0,9480 oppnådd.

Benyttes Hidden Markov Model som klassifiseringsmetode, får man en graf som er tilnærmet lik grafen i Figur 4.6. Grafen, og resultatene som ligger til grunn for denne, finnes i Vedlegg B.

Det samme eksperimentet ble også gjennomført på de to siste loggene. Resultatet fra Logg 3 var at høyest sannsynlighet for å klassifisere til riktig litelogi ble oppnådd, ved å utelate High Resolution Gamma Ray når Naive Bayes Classifier ble benyttet som klassifiseringsmetode. Klassifiseringssannsynligheten ble da tilnærmet lik 0,8313. Tas alle sensorene med, blir klassifiseringssannsynlighet på cirka 0,8285. Benyttes derimot Hidden Markov Model som klassifiseringsmetode, skiller litelogiene best ved å benytte seg av alle sensorene. En klassifiseringssannsynlighet på 0,8312 blir da oppnådd.

Ved å gjøre det samme for Logg 4, fås den høyeste klassifiseringssannsynligheten ved å utelate High Resolution Bulk Density, Thermal Neutron Porosity og High Resolution Gamma Ray når Naive Bayes Classifier benyttes som klassifiseringsmetode. Det blir oppnådd en sannsynlighet for å klassifisere litelogiene som er tilnærmet lik 0,9978. Inkluderes alle sensorene synker denne sannsynligheten til å være tilnærmet lik 0,9956.

Benyttes Hidden Markov Model oppnås best resultat om det ses bort fra de to eller tre dårligste sensorene. Det blir da oppnådd en sannsynlighet for riktig klassifisering som er tilnærmet lik 0,9978. Sannsynligheten faller til 0,9949 om alle sensorene blir inkludert.

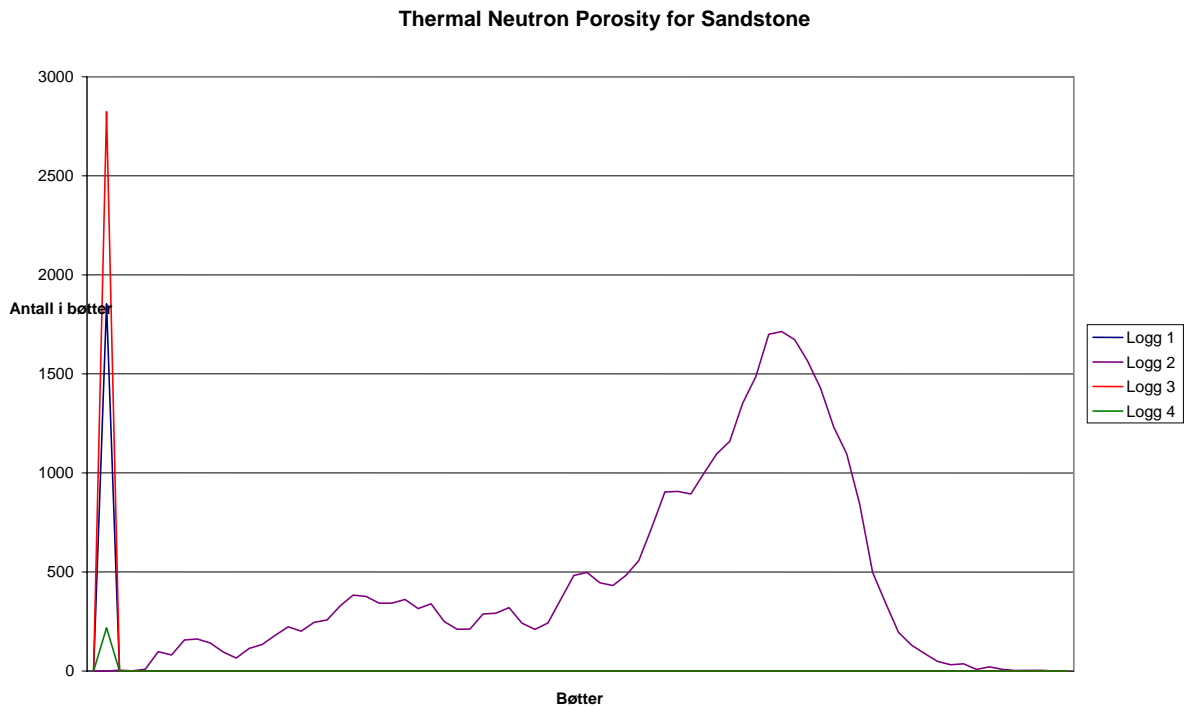
Forvirringsmatriser og grafisk fremstilling av resultatene for Logg 3 og 4 kan finnes i henholdsvis Vedlegg C og D.

4.6 Trene på tvers av loggene

Det ble også gjennomført en undersøkelse på om det vil være mulig å trene opp systemet med målingene \vec{G}_d fra en logg, for deretter å klassifisere litelogene i en annen logg ved hjelp av dette treningsgrunnlaget.

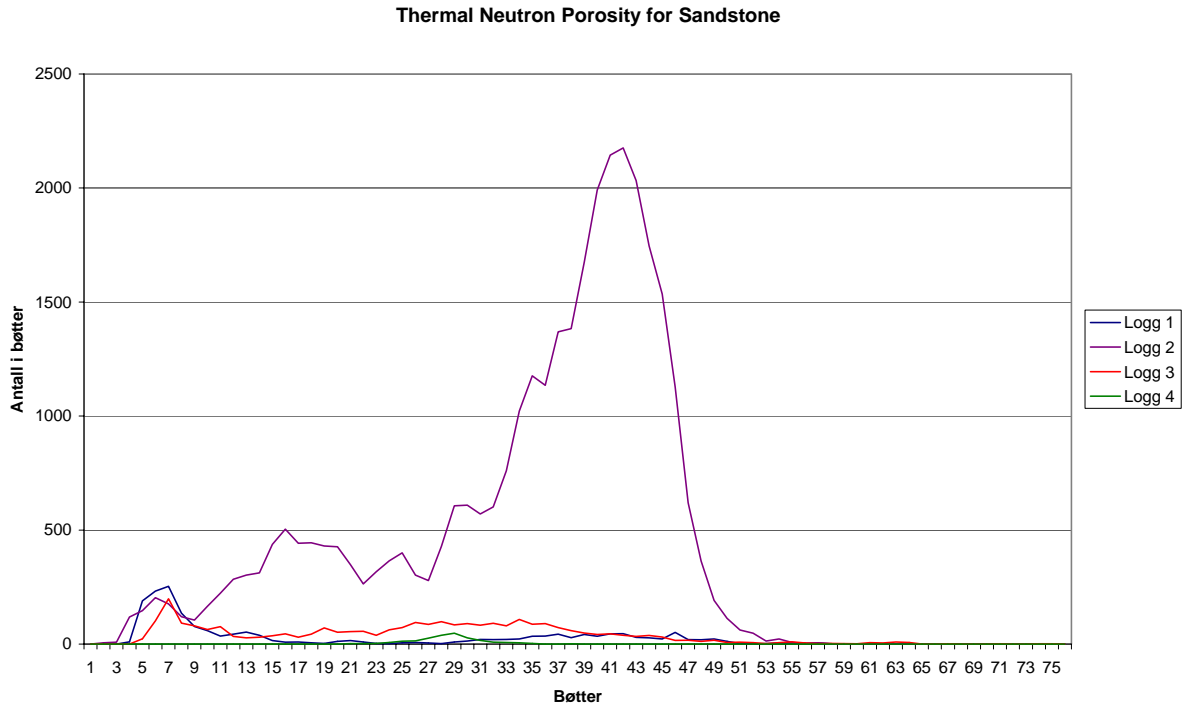
Ettersom det ikke finnes et standardformat for lagring av verdiene $G_{c,d}$ til sensorene i loggene, må det undersøkes om verdiene $G_{c,d}$ logges på samme format.

Ved å tegne verdiene $G_{c,d}$ til én litelogi fra én sensor fra de forskjellige loggene i samme figur, ser man om verdiområdet til litelogen er likt i alle loggene. Se Figur 4.7 for et eksempel på dette.



Figur 4.7: Illustrering av fordelingen til Neutron Porosity over alle loggene

Ved å se på figuren kan det se ut til at Thermal Neutron Porosity verdiene $G_{c,d}$ til Logg 2 ligger i et helt annet verdiområde enn i de tre andre loggene. Det som faktisk er tilfelle er at Logg 2 blir logget i prosent. Dette vil si at verdiene $G_{c,d}$ er 100 ganger høyere enn i de andre loggene. For å få de på samme format, slik at de kan sammenliknes, må verdiene $G_{c,d}$ divideres med 100. Ved å gjøre dette blir grafene som vist i Figur 4.8.



Figur 4.8: Illustrering av den nye fordelingen til Neutron Porosity over alle loggene.

Figuren viser at verdiene $G_{c,d}$ faktisk ligger i samme verdiområdet. De andre sensorene i loggene ble logget på samme format, og av grafene kan det se ut som om de ligger i omtrentlig det samme verdiområdet, slik som i Figur 4.8.

Vi utførte et eksperiment for å undersøke hvor godt en logg kan klassifiseres ved å benytte seg av et treningsgrunnlag beregnet fra de tre andre loggene. To av loggene ble forsøkt klassifisert på denne måten. Årsaken til at de ble valgt, ble gjort på grunnlag av hvilke litelogier og sensorer de forskjellige loggene inneholdt. Loggen som skal klassifiseres kan ikke inneholde flere litelogier eller sensorer enn det finnes i treningsgrunnlaget.

Ettersom det er stor forskjell på antallet målinger \vec{G}_d i de forskjellige loggene, ble treningen bygget opp slik at alle loggene har like stor innflytelse på treningsgrunnlaget. Om dette ikke hadde blitt gjort, ville for eksempel Sandstone i realiteten kun blitt trent opp av Logg 2 da denne har over 30 000 målinger \vec{G}_d . Logg 4 som bare har cirka 200 målinger \vec{G}_d av Sandstone, ville fått liten innflytelse på treningsgrunnlaget. Da vi gjorde dette, og undersøkte hvor godt Naive Bayes Classifier klarte å klassifisere Logg 3, fikk vi resultatene som Tabell 4.26 viser.

Tabell 4.26: Logg 1 klassifisert med Naive Bayes Classifier med nytt treningsgrunnlag

| | Sandstone | Limestone | Claystone | Anhydrite | Siltstone | Dolomite |
|-----------|-----------|-----------|-----------|-----------|-----------|----------|
| Sandstone | 477 | 1175 | 1046 | 13 | 1 | 292 |
| Limestone | 57 | 186 | 170 | 13 | 17 | 7 |
| Claystone | 60 | 642 | 11837 | 0 | 0 | 4 |

Alle: 15997
Riktig: 12500
Feil: 3497
Feilrate: 0,218603488
Klassifisering: 0,781396512

Tabell 4.26 viser at litelogiene blir klassifisert med en sannsynlighet på cirka 0,7814. Det vil i praksis si at tilnærmet 78,14 % av alle målingene \vec{c}_i blir klassifisert til riktig litelogi.

Det ble også gjennomført en undersøkelse av hvor godt Hidden Markov Model klarte å klassifisere litelogiene. Resultatene som ble oppnådd med dette eksperimentet, finnes i forvirringsmatrisen som blir vist i Tabell 4.27.

Tabell 4.27: Logg 1 klassifisert med Hidden Markov Model med nytt treningsgrunnlag

| | Sandstone | Limestone | Claystone | Anhydrite | Siltstone | Dolomite |
|-----------|-----------|-----------|-----------|-----------|-----------|----------|
| Sandstone | 490 | 1149 | 1062 | 11 | 0 | 292 |
| Limestone | 59 | 165 | 186 | 15 | 17 | 8 |
| Claystone | 81 | 564 | 11895 | 0 | 0 | 3 |

Alle: 15997
Riktig: 12550
Feil: 3447
Feilrate: 0,215477902
Klassifisering: 0,784522098

Av forvirringsmatrisen kan det leses at litelogiene blir klassifisert med en sannsynlighet som er tilnærmet lik 0,7845. Det vil si at 78,45 % av alle målingene \vec{c}_i blir klassifisert til riktig litelogi.

Et eksperiment for å undersøke om klassifiseringssannsynligheten ble påvirket av overgangsmatrisen ble utført. Dette ble gjort ved å benytte snittovergangsmatrisen, som ble laget ved å beregne snittet av verdiene for de fire spesiellagde overgangsmatrisene, i Hidden Markov Model. Resultatene som Hidden Markov Model produserte med den nye overgangsmatrisen kan ses i Tabell 4.28.

Tabell 4.28: Logg 1 klassifisert med Hidden Markov Model, nytt treningsgrunnlag og snittovergangsmatrise

| | Sandstone | Limestone | Claystone | Anhydrite | Siltstone | Dolomite |
|-----------|-----------|-----------|-----------|-----------|-----------|----------|
| Sandstone | 490 | 1149 | 1062 | 11 | 0 | 292 |
| Limestone | 61 | 167 | 184 | 13 | 17 | 8 |
| Claystone | 78 | 564 | 11898 | 0 | 0 | 3 |

Alle: 15997
Riktig: 12555
Feil: 3442
Feilrate: 0,21516534
Klassifisering: 0,78483466

Av tabellen kan man se at Hidden Markov Model, med den nye snittovergangsmatrisen, klassifiserer litelogiene med en sannsynlighet som er tilnærmet lik 0,7848. I praksis vil dette si at 78,48 % av alle målingene \vec{G}_d blir klassifisert til riktig litelogi.

Det samme eksperimentet ble utført med Logg 2. Resultatet var at alle målingene \vec{G}_d ble klassifisert til litelogien Claystone. Dette vil si at klassifiseringssannsynligheten ble 0,1719, og var den for begge klassifiseringsmetodene. Forvirringsmatrisene og nøyaktig klassifiseringssannsynlighet til denne undersøkelsen kan finnes i Vedlegg B.

5 Diskusjon av resultater

Flere forsøk og eksperimenter har blitt utført i løpet av prosjektperioden. I dette kapittelet blir eksperimentene, som ble beskrevet i Kapittel 4, drøftet. Det har blitt gjort forsøk på å klassifisere litelogiene både ved hjelp av én sensor, flere sensorer med Naive Bayes Classifier og ved hjelp av etterliggende målinger \vec{G}_a hvor Hidden Markov Model ble benyttet. Optimaliseringene som har blitt gjort for å utbedre klassifiseringsmetodene vil også bli drøftet. Det blir foretatt en vurdering av resultatene vi fikk når det ble forsøkt å klassifisere en logg ved hjelp av treningsgrunlaget fra en eller flere andre logger. Til slutt i kapittelet følger et forslag på hva som kan gjøres videre med utgangspunkt i dette prosjektet.

5.1 Én sensor

I kapittel 4.1 ble det først sett på muligheten til å klassifisere litelogiene ved hjelp av kun én sensor. Ved å se på de forskjellige sensorene i de fire loggene, så vi at flere av sensorene klarte å skille litelogiene godt alene.

Sensoren Hals Measurements / Medium Resistivity klarer å klassifisere litelogiene best i både Logg 1, 2 og 3, og i Logg 4 klassifiserer den nest best. Ut fra loggene kunne det se ut til at litelogiene kan klassifiseres godt med den ene sensoren. Sannsynligheten for riktig klassifisering av litelogiene var opp mot 0,9973. Om vi velger å klassifisere litelogiene kun ved hjelp av denne ene sensoren vil det oppstå flere problemer. Det første problemet er at det ikke logges de samme sensorene under boringen. I tilfeller der Hals Measurements / Medium Resistivity ikke finnes i loggen, kan heller ikke denne benytte til å klassifisere litelogiene.

Et annet problem er at det ikke er én sensor som er best til å klassifisere litelogiene i alle loggene. I Logg 4 klarer for eksempel Thermal Neutron Porosity å klassifisere litelogiene bedre enn Hals Measurements / Medium Resistivity, og det vil dermed være ønskelig å velge denne sensoren.

Det tredje problemet som oppstår, er at verdier $G_{c,d}$ ikke logges fra sensoren til enhver tid. I kortere eller lengre perioder kan det skje at Hals Measurements / Medium Resistivity ikke registrerer verdier $G_{c,d}$, men kun logger såkalte ”dummyverdier”. I områder der dette inntreffer, kan det ikke sies noe om hvilke litelogier det bores i.

Videre er det et problem at graden av riktig klassifisering blir for lav. Til tross for at Hals Measurements / Medium Resistivity gir best klassifiseringssannsynlighet i Logg 1, oppnår den kun en sannsynlighet for riktig klassifisering på 0,7529. Dette vil i praksis si at cirka en fjerdedel av alle verdiene $G_{c,d}$ i denne loggen vil bli klassifisert til feil litelogi.

5.2 Naive Bayes Classifier

På grunn av de fire problemene nevnt over ble det undersøkt om en høyere grad av riktig klassifisering kunne oppnås ved å benytte flere sensorer til å klassifisere litelogiene. Naive Bayes Classifier ble brukt til å undersøke dette.

Ved å prøve ut klassifiseringsmetoden på de samme loggene som nevnt tidligere, kom det fram at metoden klassifiserte litelogiene bedre enn alle de enkelte sensorene i tre av fire logger. I den siste loggen, Logg 2, var det en forskjell på 0,0494 mellom den sensoren som gav høyest grad av riktig klassifisering og Naive Bayes Classifier. Dette vil i praksis si at 4,94 prosentpoeng flere målinger \vec{G}_a blir klassifisert til feil litelogi. Noe som er verdt å merke seg i dette tilfellet, er at Naive Bayes Classifier klassifiserer mange flere målinger \vec{G}_a . Årsaken til dette er at Hals Measurements / Medium Resistivity logget ”dummyverdier”.

Ved å sammenlikne den sensoren som klassifiserte litelogiene best med resultatene som Naive Bayes Classifier oppnådde i Logg 1, oppnås det en forbedring på 11,82 prosentpoeng. Når man i tillegg tar hensyn til alle problemene som oppstår ved å benytte seg av kun én sensor, mener vi at Naive Bayes Classifier er et mye bedre alternativ.

5.3 Hidden Markov Model

Det ble undersøkt om det fantes en korrelasjon mellom målinger \vec{G}_a utført på flere boreddybder. Metoden Hidden Markov Model undersøker om det kan oppnås en bedre klassifisering ved å benytte seg av denne korrelasjonen.

Vi benyttet de samme loggene med Hidden Markov Model som klassifiseringsmetode. Når vi sammenliknet denne metoden med resultatene fra Naive Bayes Classifier, oppnådde vi en høyere sannsynlighet for riktig klassifisering i tre av de fire loggene.

Logg 4 klassifiserte litelogiene dårligere når Hidden Markov Model ble benyttet som klassifiseringsmetode, og forskjellen i klassifiseringssannsynligheten var på 0,0007. Det vil i praksis si at det skjer en økning på 0,07 prosentpoeng av antall målinger \vec{G}_a som blir klassifisert til feil litelogi.

Ved å benytte Hidden Markov Model som klassifiseringsmetode på de tre andre loggene, vil den høyeste økningen i riktig klassifisering oppnås i Logg 1. I denne loggen oppnås en økning i klassifiseringssannsynlighet på 0,0092. Det vil si en økning på 0,92 prosentpoeng av antall målinger \vec{G}_a som blir klassifisert til riktig litelogi.

Selv om forskjellen ikke var av betydelig størrelse, oppnås det i de fleste tilfeller en høyere grad av riktig klassifisering ved å benytte Hidden Markov Model. Ut i fra dette leser vi at det finnes en korrelasjon mellom målingene \vec{G}_a , og at denne kan benyttes til å utvikle en bedre klassifikator.

Det ble også undersøkt om en generell snittovergangsmatrise kan benyttes, eller om dette vil påvirke resultatene til Hidden Markov Model. Se Kapittel 4.4 for beskrivelse av overgangsmatriser. Ved å benytte Hidden Markov Model med en snittovergangsmatrise, som ble laget som et snitt av de spesiellagede overgangsmatrisene, ble det observert at resultatene ikke ble påvirket i noen merkbare grad.

5.4 Utbedringer av klassifikatorene

Ved å se på litelogiprofilene kunne det virke som om sedimenter ville kunne påvirke klassifiseringen av litelogiene, ved at de førte til en plutselig sporadisk økning i verdiene $G_{c,d}$ målt fra sensorene. Se Kapittel 2.3 for beskrivelse av hva et sediment er. Vi undersøkte derfor om klassifiseringen kunne bedres ved å kutte ut 1 % av de såkalte uteliggerne. Se Kapittel 4.5 for beskrivelse av uteliggere. Ved å gjøre dette ble resultatene at to av loggene ble klassifisert bedre, mens de to andre ble klassifisert dårligere.

Ut fra disse resultatene kunne det tyde på at det var for mye å fjerne 1 % av uteliggerne i noen tilfeller. Vi forsøkte derfor å senke grensen på hvor stor andel av uteliggerne som ble fjernet, for å oppnå et positivt resultat i alle tilfeller. Med dette som utgangspunkt forsøkte vi derfor å fjerne 1 ‰ av uteliggerne. Resultatet av dette var at klassifiseringssannsynligheten til Naive Bayes Classifier og Hidden Markov Model steg eller forble uforandret i alle de fire loggene. Ved å fjerne 1 ‰ av uteliggerne kan en forbedring av klassifiseringssannsynligheten på 0,0095 oppnås, dersom Naive Bayes Classifier blir benyttet som klassifiseringsmetode. I praksis vil dette si at litelogiene blir klassifisert bedre med 0,95 prosentpoeng. Ved å benytte Hidden Markov Model som klassifiseringsmetode, oppnås en økning i klassifiseringssannsynligheten på cirka 0,0082.

Selv om det ikke oppnås en stor forbedring av klassifiseringssannsynligheten ved å fjerne 1 ‰ av uteliggerne, vil vi allikevel anbefale dette da et positivt eller like bra resultat blir oppnådd i alle loggene.

En annen detalj som vi merket oss, når vi undersøkte hvor godt litelogiene ble klassifisert ved hjelp av én sensor, var at enkelte sensorer klassifiserte litelogiene dårlig. Om man tar sensoren High Resolution Gamma Ray som eksempel, ser man at denne klassifiserer dårligst i Logg 2 og 4, og i Logg 3 klassifiserer den nest dårligst. I Logg 1 er sensoren den fjerde dårligste sensoren til å skille litelogiene.

Det ble derfor gjort en undersøkelse på om klassifiseringen blir bedre ved å fjerne en eller flere av de sensorene som klassifiserte dårligst. Ved å undersøke dette på Logg 2, fikk vi høyest grad av riktig klassifisering ved å unngå å benytte Medium Shallow Resistivity. Undersøkes derimot Logg 1, vil loggen få en vesentlig lavere klassifiseringssannsynlighet dersom Medium Shallow Resistivity blir utelatt. Blir Naive Bayes Classifier benyttet som klassifiseringsmetode på Logg 3, vil man få en bedre høyere grad av riktig klassifisering ved å utelate High Resolution Gamma Ray. Benyttes Hidden Markov Model som klassifiseringsmetode, vil en bedre klassifiseringssannsynlighet oppnås ved å inkludere denne sensoren.

I Logg 2 kan det ses at Medium Shallow Resistivity skiller litelogiene dårlig. Ved kun å benytte denne sensoren klassifiseres 34,04 % av målingene \vec{g}_d til riktig litelogi. Benyttes den samme sensoren i Logg 1, klassifiseres målingene \vec{g}_d til riktig litelogi i 74,49 % av tilfellene. Dette tilsier at sensoren er nest best til å klassifisere litelogiene i denne loggen. Grunnet de store variasjonene i hvor godt sensorene klassifiserer litelogiene i de forskjellige loggene, kan det ikke trekkes en konklusjon at det vil være en fordel å utelate noen av sensorene.

5.5 Trene på tvers av logger

Det ble gjort et eksperiment hvor det ble undersøkt hvor godt en logg ble klassifisert når de tre andre loggene ble brukt som treningsgrunnlag. Dette ble gjort for å undersøke om verdiområdene, som kjennetegner hver litelogi, forandrer seg avhengig av hvor det bores.

Resultatene vi fikk, viser at klassifiseringen av litelogiene ble på langt nær så god som når treningsgrunnlaget kom fra samme borehull. Årsaken til dette er at verdiområdet til sensorene forandrer seg, og er dermed ulike i de forskjellige loggene. Dette fører til store standardavvik når de blir benyttet som treningsgrunnlag. En konsekvens av dette var at flere målinger \vec{G}_i ble klassifisert feil, noe som igjen førte til at klassifiseringen ble dårligere.

I tillegg ble det foretatt en undersøkelse hvor vi så om verdiområdet som kjennetenger litelogiene forandrer seg avhengig av boreddybde. Det viste seg at så ikke var tilfelle.

5.6 Videre arbeid

Programmet som har blitt utviklet i dette prosjektet vil ikke kunne tas i bruk slik det er nå uten modifikasjoner. Det ble kun laget for å undersøke hvor godt Naive Bayes Classifier og Hidden Markov Model egnest seg som automatisk loggtolker. Om det ønskes å utvikle et slikt system som skal kunne tas i bruk på en boreplattform, er det flere områder som bør undersøkes.

Det første problemet som bør undersøkes nærmere er shale-baseline-shift. Dette er et spesielt tilfelle som bare kan skje for sensoren High Resolution Gamma Ray, og fører til at gammaspekteret blir flyttet noen oktaver opp eller ned. Dette kommer av at enten boreslammet inneholder kalium eller at kalibreringen av gammasensoren ikke er helt rett. Det som må undersøkes er hvor ofte disse skiftene inntreffer og eventuelt oppdatere systemet så det klarer å oppdage slike skifter.

Det bør også undersøkes om det er mulig å oppdage skift i verdiområdet som representerer litelogiene fra borehull til borehull. Dersom man klarer dette, kan klassifiseringsmetoden trenes opp generelt.

For å få en så sikker klassifisering av litelogiene som mulig, bør også andre klassifiseringsmetoder undersøkes. Ettersom man aldri vil kunne lage et klassifiseringssystem som klassifiserer riktig i alle tilfeller, kan det være interessant å undersøke om en høyere klassifiseringssannsynlighet kan oppnås ved å benytte flere metoder sammen.

Det bør også innhentes mer data for flere litelogier for å undersøke hvor godt man klarer å klassifisere de med hjelp av Naive Bayes Classifier og Hidden Markov Model.

6 Konklusjon

I denne rapporten er hovedvekten av arbeidet lagt på å undersøke hvor godt metodene, Naive Bayes Classifier og Hidden Markov Model egner seg til en litelogianalyse.

Vi undersøkte hvor godt litelogiene ble klassifisert ved kun å benytte én sensor. Undersøkelsen viste at det var stor variasjon i klassifiseringssannsynligheten til sensorene avhengig av hvilken logg som ble benyttet. Hovedproblemet med å kun benytte én sensor er at det ikke er sikkert at en sensor vil finnes i alle loggene. Dette fører til at denne løsningen egner seg dårlig til å skille litelogiene.

Det ble undersøkt hvor godt litelogiene ble skilt ved hjelp av Naive Bayes Classifier som benytter flere sensorer. Metoden ble sammenliknet med resultatene ved å klassifisere med én enkelt sensor, og det viste seg at Naive Bayes Classifier klassifiserte bedre i tre av fire logger. Denne klassifiseringsmetoden klarte å klassifisere loggene forholdsvis bra, og en klassifiseringssannsynlighet på 0,8285 eller høyere ble oppnådd.

Videre undersøkte vi i hvilken grad det er korrelasjon mellom etterfølgende målinger \vec{G}_d . Ved å benytte Hidden Markov Model som klassifiseringsmetode, ble det undersøkt om korrelasjonen kan brukes til å øke klassifiseringssannsynligheten. Sammenliknes resultatene med Naive Bayes Classifier, så vi at Hidden Markov Model klassifiserte tre av de fire loggene bedre. Resultatene tyder på at det finnes en slik korrelasjon mellom målingene \vec{G}_d . Ved å bruke Hidden Markov Model som klassifiseringsmetode, ble en klassifiseringssannsynlighet på 0,8312 eller høyere oppnådd i de fire loggene.

Det ble foretatt en undersøkelse av hvor stor innvirkning det hadde på klassifiseringssannsynligheten til Hidden Markov Model at det ble benyttet spesiallagde overgangsmatriser. Ved å benytte seg av en snittovergangsmatrise, som ble laget på grunnlag av alle de spesiallagde overgangsmatrisene, så vi at dette hadde liten eller ingen innvirkning på klassifiseringssannsynligheten.

Vi så også på optimaliseringer for å utbedre klassifiseringsmetodene. Om 1 % av uteliggerne ble fjernet, oppnås en forbedring av klassifiseringssannsynligheten. Årsaken til dette er at innslaget av sedimenter i litelogiene fører til feilklassifiseringer. Noen av disse feilene blir rettet ved å fjerne 1 % av de høyeste verdiene $G_{c,d}$ til hver sensor i hver litelogi.

Klassifiseringen vil ikke nødvendigvis bli bedre ved å utelate de sensorene som klassifiserer litelogiene dårligst. I noen tilfeller kan det slå positivt ut når disse sensorene blir utelatt, men i andre logger kan det føre til at klassifiseringen blir dårligere. Årsaken til dette er den store variasjonen i hvor godt litelogiene kan bli skilt i de forskjellige loggene av én sensor. Dette fører til at det ikke kan slås fast at én sensor alltid vil klassifiserer litelogiene dårligst, og på grunn av dette, anbefaler vi at alle tilgjengelige sensorer bør benyttes.

Det ble undersøkt hvor godt litelogiene i en logg ble klassifisert, ved å benytte målinger \vec{G}_d fra andre logger som treningsgrunnlag. Eksperimentet ble gjennomført for å undersøke om verdiområdene, som kjennetegner hver litelogi, forandrer seg avhengig av hvor det bores. Resultatet av dette eksperimentet var at graden av riktig klassifisering sank betraktelig. Årsaken til dette er at verdiområdet til sensorene forandrer seg, noe som fører til store standardavvik. En konsekvens av dette kan være at flere målinger \vec{G}_d blir feilklassifisert, noe som igjen fører til at klassifiseringssannsynligheten blir dårligere. I tillegg ble det foretatt en undersøkelse hvor vi så om verdiområdet som kjennetegner litelogiene forandrer seg avhengig av boreddybde. Det viste seg at så ikke var tilfelle.

Ut fra resultatene i denne rapporten ser det ut til at Hidden Markov Model bør bli benyttet som klassifiseringsmetode. Alle tilgjengelige sensorer bør benyttes, og 1 % av uteliggerne bør fjernes. Systemet må trenes opp med data fra samme borehull for å oppnå høyest grad av riktig klassifisering. Om resultatene ønskes i "sanntid" mens boringen pågår, vil Naive Bayes Classifier være et godt alternativ til Hidden Markov Model. Eventuelt kan Hidden Markov Model konfigurere slik at den tilnærmet klassifiserer i "sanntid" ved for eksempel å benytte kun fem etterliggende målinger \vec{G}_d .

Referanser

- [1] Richard O. Duda, Peter E. Hart, David G. Stork (2001). *Pattern Classification, Second Edition*. US: John Wiley & Sons, LTD.
- [2] Branden Fitelson (2001). *Studies in Bayesian Confirmation Theory*.
- [3] <http://plato.stanford.edu/entries/bayes-theorem/>
Hentet: 14.04.05.
- [4] I. Rish. *An empirical study of the naive Bayes classifier*.
- [5] Lawrence R. Rabiner (1989). *A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition*.
- [6] Phil Blunsom (2004). *Hidden Markov Models*.
- [7] Takao Aizawa, Tsutomu Hashimoto, Nirou Okamoto and Hiroshi Karshima (2004). *Lithology Analysis Using Neutron – Gamma Logging*.
- [8] Hsien-cheng Chang, Hui-Chuan Chen, and Jen-Ho Fang (1997). *Lithology Determination from Well Logs with Fuzzy Associative Memory Neural Network*.
- [9] LogTrans – CMTE Software for Automated Interpretation of Borehole Geophysical Logs
<http://www.cmte.org.au/research/logtrans.html>
Hentet 11.02.04.
- [10] <http://www.glossary.oilfield.slb.com>
Hentet: 25.04.05.
- [11] <http://www.connect.slb.com/Docs/connect/reference/Chartbook/>
Hentet 25.04.05.
- [12] H. Delius, S. Hunze, R. Pechinig, A. Bartetzko, J. Wohlenberg and Shipboard Scientific Party. *Shorebased Interpretation of Downhole Measurements at Sites 1065, 1068 and 1069*.

Vedlegg

Her følger alle resultatene som vi har produsert i løpet av prosjektperioden. Disse er ment som en bakgrunnsinformasjon for å kunne sette seg inn i alle klassifiseringssannsynlighetene i alle loggene. Det er på grunnlag av disse resultatene at konklusjonene har blitt trukket.

Vedlegg A – Forvirringsmatriser for Logg 1

Vedlegg B – Forvirringsmatriser for Logg 2

Vedlegg C – Forvirringsmatriser for Logg 3

Vedlegg D – Forvirringsmatriser for Logg 4

Vedlegg A – Forvirringsmatriser for Logg 1

Enkelt sensorer

Deep Resistivity

| | Sandstone | Limestone | Claystone | Siltstone |
|-----------|-----------|-----------|-----------|-----------|
| Sandstone | 1183 | 40 | 723 | 0 |
| Limestone | 55 | 8 | 142 | 0 |
| Claystone | 17 | 23 | 260 | 0 |
| Siltstone | 0 | 0 | 0 | 0 |

Alle: 2451
Riktig: 1451
Feil: 1000
Feilrate: 0,40799674
Klassifisering: 0,59200326

High Resolution Bulk Density

| | Sandstone | Limestone | Claystone | Siltstone |
|-----------|-----------|-----------|-----------|-----------|
| Sandstone | 1284 | 589 | 0 | 0 |
| Limestone | 44 | 161 | 0 | 0 |
| Claystone | 93 | 169 | 38 | 0 |
| Siltstone | 0 | 0 | 0 | 0 |

Alle: 2378
Riktig: 1483
Feil: 895
Feilrate: 0,37636669
Klassifisering: 0,62363331

High Resolution Gamma Ray

| | Sandstone | Limestone | Claystone | Siltstone |
|-----------|-----------|-----------|-----------|-----------|
| Sandstone | 785 | 393 | 1011 | 132 |
| Limestone | 38 | 33 | 67 | 67 |
| Claystone | 1449 | 818 | 7778 | 966 |
| Siltstone | 0 | 4 | 0 | 91 |

Alle: 13632
Riktig: 8687
Feil: 4945
Feilrate: 0,36274941
Klassifisering: 0,63725059

Medium Resistivity / Mud Resistivity from HALS Measurements

| | Sandstone | Limestone | Claystone | Siltstone |
|-----------|-----------|-----------|-----------|-----------|
| Sandstone | 1125 | 178 | 952 | 66 |
| Limestone | 40 | 54 | 14 | 97 |
| Claystone | 0 | 59 | 8980 | 1936 |
| Siltstone | 0 | 0 | 17 | 78 |

Alle: 13596
Riktig: 10237
Feil: 3359
Feilrate: 0,24705796
Klassifisering: 0,75294204

Medium Shallow Resistivity

| | Sandstone | Limestone | Claystone | Siltstone |
|-----------|-----------|-----------|-----------|-----------|
| Sandstone | 949 | 319 | 855 | 198 |
| Limestone | 5 | 78 | 54 | 68 |
| Claystone | 0 | 44 | 9012 | 1919 |
| Siltstone | 0 | 0 | 6 | 89 |

Alle: 13596
Riktig: 10128
Feil: 3468
Feilrate: 0,25507502
Klassifisering: 0,74492498

Thermal Neutron Porosity

| | Sandstone | Limestone | Claystone |
|-----------|-----------|-----------|-----------|
| Sandstone | 1041 | 186 | 626 |
| Limestone | 0 | 172 | 33 |
| Claystone | 10 | 46 | 92 |

Alle: 2206
Riktig: 1305
Feil: 901
Feilrate: 0,40843155
Klassifisering: 0,59156845

Naive Bayes Classifier

| | Sandstone | Limestone | Claystone | Siltstone |
|-----------|-----------|-----------|-----------|-----------|
| Sandstone | 1169 | 111 | 956 | 88 |
| Limestone | 41 | 80 | 84 | 0 |
| Claystone | 10 | 44 | 10537 | 423 |
| Siltstone | 0 | 0 | 1 | 94 |

Alle: 13638
Riktig: 11880
Feil: 1758
Feilrate: 0,12890453
Klassifisering: 0,87109547

Hidden Markov Model

| | Sandstone | Limestone | Claystone | Siltstone |
|-----------|-----------|-----------|-----------|-----------|
| Sandstone | 1173 | 106 | 980 | 65 |
| Limestone | 41 | 79 | 85 | 0 |
| Claystone | 11 | 40 | 10659 | 304 |
| Siltstone | 0 | 0 | 0 | 95 |

Alle: 13638
Riktig: 12006
Feil: 1632
Feilrate: 0,11966564
Klassifisering: 0,88033436

Hidden Markov Model med snittovergangsmatrise

| | Sandstone | Limestone | Claystone | Siltstone |
|-----------|-----------|-----------|-----------|-----------|
| Sandstone | 1173 | 107 | 979 | 65 |
| Limestone | 41 | 80 | 84 | 0 |
| Claystone | 11 | 39 | 10661 | 303 |
| Siltstone | 0 | 0 | 0 | 95 |

Alle: 13638
Riktig: 12009
Feil: 1629
Feilrate: 0,11944567
Klassifisering: 0,88055433

Naive Bayes Classifier når 1 % av uteliggerne er fjernet

| | Sandstone | Limestone | Claystone | Siltstone |
|-----------|-----------|-----------|-----------|-----------|
| Sandstone | 1169 | 158 | 922 | 74 |
| Limestone | 39 | 114 | 52 | 0 |
| Claystone | 46 | 94 | 10479 | 395 |
| Siltstone | 0 | 0 | 0 | 95 |

Alle: 13637
Riktig: 11857
Feil: 1780
Feilrate: 0,13052724
Klassifisering: 0,86947276

Hidden Markov Model når 1 % av uteliggerne er fjernet

| | Sandstone | Limestone | Claystone | Siltstone |
|-----------|-----------|-----------|-----------|-----------|
| Sandstone | 1172 | 145 | 926 | 80 |
| Limestone | 41 | 109 | 55 | 0 |
| Claystone | 40 | 85 | 10575 | 314 |
| Siltstone | 0 | 0 | 0 | 95 |

Alle: 13637
Riktig: 11951
Feil: 1686
Feilrate: 0,12363423
Klassifisering: 0,87636577

Naive Bayes Classifier når 1 % av uteliggerne er fjernet

| | Sandstone | Limestone | Claystone | Siltstone |
|-----------|-----------|-----------|-----------|-----------|
| Sandstone | 1175 | 141 | 926 | 81 |
| Limestone | 41 | 102 | 62 | 0 |
| Claystone | 12 | 59 | 10578 | 365 |
| Siltstone | 0 | 0 | 1 | 94 |

Alle: 13637
Riktig: 11949
Feil: 1688
Feilrate: 0,12378089
Klassifisering: 0,87621911

Hidden Markov Model når 1 % av uteliggerne er fjernet

| | Sandstone | Limestone | Claystone | Siltstone |
|-----------|-----------|-----------|-----------|-----------|
| Sandstone | 1178 | 130 | 935 | 80 |
| Limestone | 41 | 102 | 62 | 0 |
| Claystone | 12 | 56 | 10675 | 271 |
| Siltstone | 0 | 0 | 0 | 95 |

Alle: 13637
Riktig: 12050
Feil: 1587
Feilrate: 0,11637457
Klassifisering: 0,88362543

Antall sensorer man bør benytte på Logg 1 med Naive Bayes Classifier

Medium Resistivity / Mud Resistivity from HALS Measurements

| | Sandstone | Limestone | Claystone | Siltstone |
|-----------|-----------|-----------|-----------|-----------|
| Sandstone | 1125 | 178 | 952 | 66 |
| Limestone | 40 | 54 | 14 | 97 |
| Claystone | 0 | 59 | 8980 | 1936 |
| Siltstone | 0 | 0 | 17 | 78 |

Alle: 13596
Riktig: 10237
Feil: 3359
Feilrate: 0,24705796
Klassifisering: 0,75294204

Medium Resistivity / Mud Resistivity from HALS Measurements, Deep Resistivity

| | Sandstone | Limestone | Claystone | Siltstone |
|-----------|-----------|-----------|-----------|-----------|
| Sandstone | 1163 | 141 | 977 | 41 |
| Limestone | 50 | 46 | 109 | 0 |
| Claystone | 2 | 58 | 9001 | 1914 |
| Siltstone | 0 | 0 | 17 | 78 |

Alle: 13597
Riktig: 10288
Feil: 3309
Feilrate: 0,24336251
Klassifisering: 0,75663749

Medium Resistivity / Mud Resistivity from HALS Measurements, Deep Resistivity, Thermal Neutron Porosity

| | Sandstone | Limestone | Claystone | Siltstone |
|-----------|-----------|-----------|-----------|-----------|
| Sandstone | 1172 | 110 | 1001 | 40 |
| Limestone | 40 | 70 | 95 | 0 |
| Claystone | 7 | 52 | 9002 | 1914 |
| Siltstone | 0 | 0 | 17 | 78 |

Alle: 13598
Riktig: 10322
Feil: 3276
Feilrate: 0,24091778
Klassifisering: 0,75908222

Medium Resistivity / Mud Resistivity from HALS Measurements, Deep Resistivity, Thermal Neutron Porosity, High Resolution Bulk Density

| | Sandstone | Limestone | Claystone | Siltstone |
|-----------|-----------|-----------|-----------|-----------|
| Sandstone | 1191 | 92 | 1000 | 40 |
| Limestone | 41 | 78 | 86 | 0 |
| Claystone | 40 | 19 | 9002 | 1914 |
| Siltstone | 0 | 0 | 17 | 78 |

Alle: 13598
Riktig: 10349
Feil: 3249
Feilrate: 0,2389322
Klassifisering: 0,7610678

Medium Resistivity / Mud Resistivity from HALS Measurements, Deep Resistivity, Thermal Neutron Porosity, High Resolution Bulk Density, High Resolution Gamma Ray

| | Sandstone | Limestone | Claystone | Siltstone |
|-----------|-----------|-----------|-----------|-----------|
| Sandstone | 1199 | 194 | 802 | 128 |
| Limestone | 42 | 91 | 72 | |
| Claystone | 40 | 64 | 10119 | 791 |
| Siltstone | 0 | 0 | 3 | 92 |

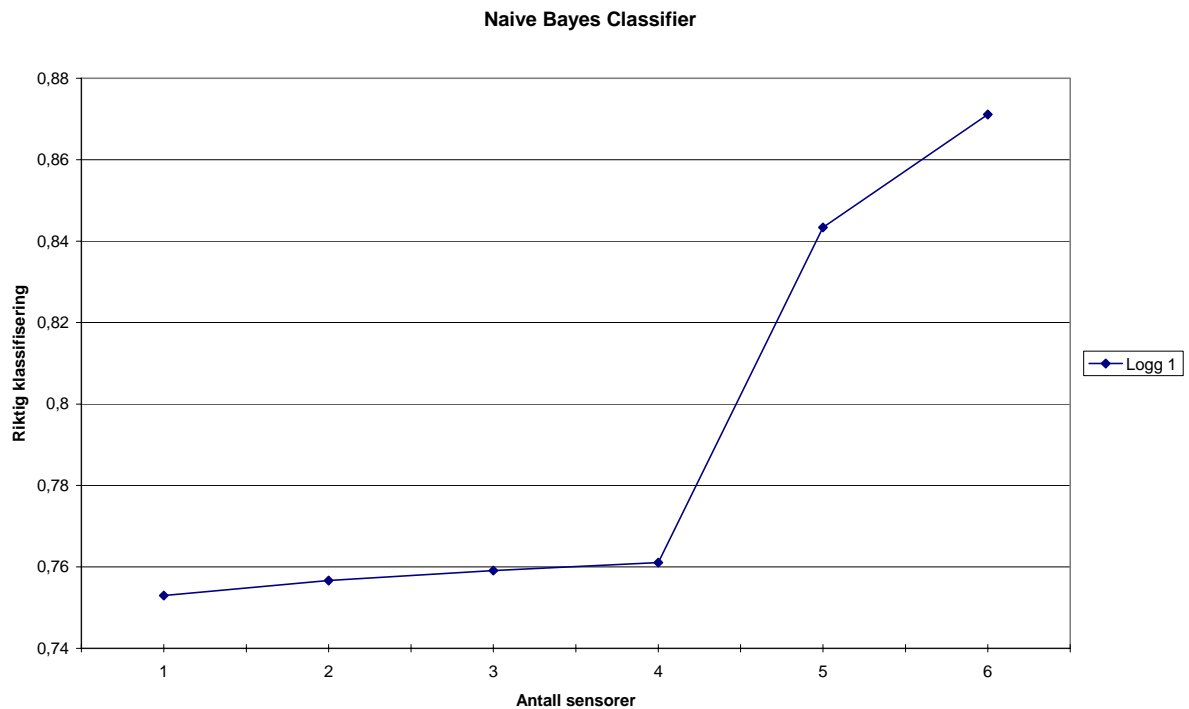
Alle: 13637
Riktig: 11501
Feil: 2136
Feilrate: 0,15663269
Klassifisering: 0,84336731

Medium Resistivity / Mud Resistivity from HALS Measurements, Deep Resistivity, Thermal Neutron Porosity, High Resolution Bulk Density, High Resolution Gamma Ray, Medium Shallow Resistivity (lik Naive Bayes Classifier over)

| | Sandstone | Limestone | Claystone | Siltstone |
|-----------|-----------|-----------|-----------|-----------|
| Sandstone | 1169 | 111 | 956 | 88 |
| Limestone | 41 | 80 | 84 | 0 |
| Claystone | 10 | 44 | 10537 | 423 |
| Siltstone | 0 | 0 | 1 | 94 |

Alle: 13638
Riktig: 11880
Feil: 1758
Feilrate: 0,12890453
Klassifisering: 0,87109547

Grafisk fremstilling av resultatene



Antall sensorer man bør benytte på Logg 1 med Hidden Markov Model

Medium Resistivity / Mud Resistivity from HALS Measurements

| | Sandstone | Limestone | Claystone | Siltstone |
|-----------|-----------|-----------|-----------|-----------|
| Sandstone | 1146 | 157 | 990 | 28 |
| Limestone | 55 | 36 | 41 | 73 |
| Claystone | 2 | 38 | 9100 | 1835 |
| Siltstone | 0 | 0 | 17 | 78 |

Alle: 13596
Riktig: 10360
Feil: 3236
Feilrate: 0,23801118
Klassifisering: 0,76198882

Medium Resistivity / Mud Resistivity from HALS Measurements, Deep Resistivity

| | Sandstone | Limestone | Claystone | Siltstone |
|-----------|-----------|-----------|-----------|-----------|
| Sandstone | 1169 | 135 | 997 | 21 |
| Limestone | 55 | 37 | 113 | 0 |
| Claystone | 3 | 37 | 9120 | 1815 |
| Siltstone | 0 | 0 | 17 | 78 |

Alle: 13597
Riktig: 10404
Feil: 3193
Feilrate: 0,23483121
Klassifisering: 0,76516879

Medium Resistivity / Mud Resistivity from HALS Measurements, Deep Resistivity, Thermal Neutron Porosity

| | Sandstone | Limestone | Claystone | Siltstone |
|-----------|-----------|-----------|-----------|-----------|
| Sandstone | 1179 | 101 | 1022 | 21 |
| Limestone | 42 | 61 | 102 | 0 |
| Claystone | 7 | 33 | 9120 | 1815 |
| Siltstone | 0 | 0 | 17 | 78 |

Alle: 13598
Riktig: 10438
Feil: 3160
Feilrate: 0,23238712
Klassifisering: 0,76761288

Medium Resistivity / Mud Resistivity from HALS Measurements, Deep Resistivity, Thermal Neutron Porosity, High Resolution Bulk Density

| | Sandstone | Limestone | Claystone | Siltstone |
|-----------|-----------|-----------|-----------|-----------|
| Sandstone | 1177 | 103 | 1022 | 21 |
| Limestone | 42 | 76 | 87 | 0 |
| Claystone | 40 | 0 | 9120 | 1815 |
| Siltstone | 0 | 0 | 17 | 78 |

Alle: 13598
Riktig: 10451
Feil: 3147
Feilrate: 0,23143109
Klassifisering: 0,76856891

Medium Resistivity / Mud Resistivity from HALS Measurements, Deep Resistivity, Thermal Neutron Porosity, High Resolution Bulk Density, High Resolution Gamma Ray

| | Sandstone | Limestone | Claystone | Siltstone |
|-----------|-----------|-----------|-----------|-----------|
| Sandstone | 1176 | 186 | 832 | 129 |
| Limestone | 43 | 82 | 80 | 0 |
| Claystone | 40 | 87 | 10447 | 440 |
| Siltstone | 0 | 0 | 1 | 94 |

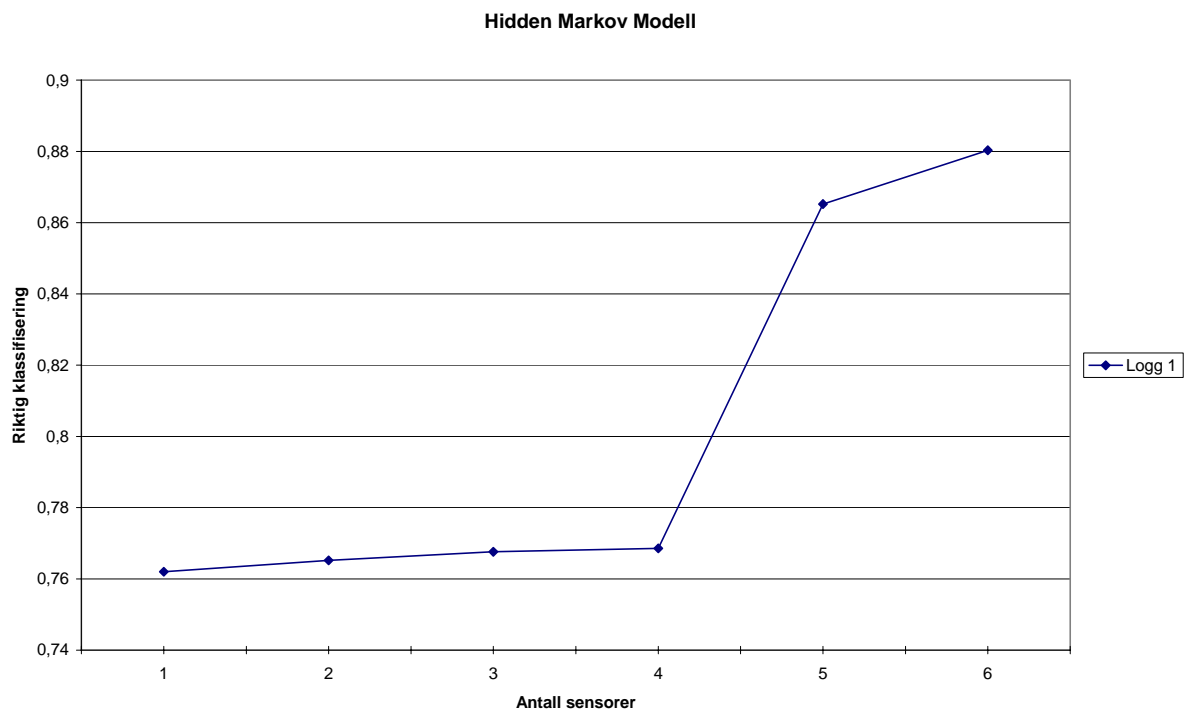
Alle: 13637
Riktig: 11799
Feil: 1838
Feilrate: 0,13478038
Klassifisering: 0,86521962

Medium Resistivity / Mud Resistivity from HALS Measurements, Deep Resistivity, Thermal Neutron Porosity, High Resolution Bulk Density, High Resolution Gamma Ray, Medium Shallow Resistivity (lik Hidden Markov Model over)

| | Sandstone | Limestone | Claystone | Siltstone |
|-----------|-----------|-----------|-----------|-----------|
| Sandstone | 1173 | 106 | 980 | 65 |
| Limestone | 41 | 79 | 85 | 0 |
| Claystone | 11 | 40 | 10659 | 304 |
| Siltstone | 0 | 0 | 0 | 95 |

Alle: 13638
Riktig: 12006
Feil: 1632
Feilrate: 0,11966564
Klassifisering: 0,88033436

Grafisk fremstilling av resultatene.



Vedlegg B – Forvirringsmatriser for Logg 2

Enkelt sensorer

Deep Resistivity

| | Sandstone | Claystone |
|-----------|-----------|-----------|
| Sandstone | 29695 | 961 |
| Claystone | 23 | 387 |

Alle: 31066
Riktig: 30082
Feil: 984
Feilrate: 0,0316745
Klassifisering: 0,9683255

High Resolution Bulk Density

| | Sandstone | Claystone |
|-----------|-----------|-----------|
| Sandstone | 21416 | 11314 |
| Claystone | 33 | 351 |

Alle: 33114
Riktig: 21767
Feil: 11347
Feilrate: 0,34266473
Klassifisering: 0,65733527

High Resolution Gamma Ray

| | Sandstone | Claystone |
|-----------|-----------|-----------|
| Sandstone | 22627 | 10274 |
| Claystone | 3561 | 3284 |

Alle: 39746
Riktig: 25911
Feil: 13835
Feilrate: 0,34808534
Klassifisering: 0,65191466

Medium Resistivity / Mud Resistivity from HALS Measurements

| | Sandstone | Claystone |
|-----------|-----------|-----------|
| Sandstone | 30642 | 14 |
| Claystone | 86 | 6746 |

Alle: 37488
Riktig: 37388
Feil: 100
Feilrate: 0,00266752
Klassifisering: 0,99733248

Medium Shallow Resistivity

| | Sandstone | Claystone |
|-----------|-----------|-----------|
| Sandstone | 4872 | 10165 |
| Claystone | 24 | 386 |

Alle: 15447
Riktig: 5258
Feil: 10189
Feilrate: 0,65961028
Klassifisering: 0,34038972

Shallow Resistivity

| | Sandstone | Claystone |
|-----------|-----------|-----------|
| Sandstone | 26630 | 4026 |
| Claystone | 62 | 6770 |

Alle: 37488
Riktig: 33400
Feil: 4088
Feilrate: 0,10904823
Klassifisering: 0,89095177

Thermal Neutron Porosity

| | Sandstone | Claystone |
|-----------|-----------|-----------|
| Sandstone | 26019 | 6644 |
| Claystone | 39 | 371 |

Alle: 33073
Riktig: 26390
Feil: 6683
Feilrate: 0,20206815
Klassifisering: 0,79793185

Naive Bayes Classifier

| | Sandstone | Claystone | Limestone |
|-----------|-----------|-----------|-----------|
| Sandstone | 30993 | 1612 | 375 |
| Claystone | 82 | 6764 | 0 |

Alle: 39826
Riktig: 37757
Feil: 2069
Feilrate: 0,05195099
Klassifisering: 0,94804901

Hidden Markov Model

| | Sandstone | Claystone | Limestone |
|-----------|-----------|-----------|-----------|
| Sandstone | 31070 | 1548 | 362 |
| Claystone | 75 | 6771 | 0 |

Alle: 39826
Riktig: 37841
Feil: 1985
Feilrate: 0,04984181
Klassifisering: 0,95015819

Hidden Markov Model med snittovergangsmatrise

| | Sandstone | Claystone | Limestone | Anhydrite | Dolomite |
|-----------|-----------|-----------|-----------|-----------|----------|
| Sandstone | 31014 | 1597 | 184 | 10 | 175 |
| Claystone | 75 | 6771 | 0 | 0 | 0 |

Alle: 39826
Riktig: 37785
Feil: 2041
Feilrate: 0,051247928
Klassifisering: 0,948752072

Naive Bayes Classifier når 1 % av uteliggerne er fjernet

| | Sandstone | Claystone | Limestone |
|-----------|-----------|-----------|-----------|
| Sandstone | 32235 | 32 | 710 |
| Claystone | 159 | 6687 | 0 |

Alle: 39823
Riktig: 38922
Feil: 901
Feilrate: 0,02262512
Klassifisering: 0,97737488

Hidden Markov Model når 1 % av uteliggerne er fjernet

| | Sandstone | Claystone | Limestone |
|-----------|-----------|-----------|-----------|
| Sandstone | 32221 | 23 | 733 |
| Claystone | 143 | 6703 | 0 |

Alle: 39823
Riktig: 38924
Feil: 899
Feilrate: 0,02257489
Klassifisering: 0,97742511

Naive Bayes Classifier når 1 % av uteliggerne er fjernet

| | Sandstone | Claystone | Limestone |
|-----------|-----------|-----------|-----------|
| Sandstone | 31387 | 942 | 650 |
| Claystone | 100 | 6746 | 0 |

Alle: 39825
Riktig: 38133
Feil: 1692
Feilrate: 0,04248588
Klassifisering: 0,95751412

Hidden Markov Model når 1 % av uteliggerne er fjernet

| | Sandstone | Claystone | Limestone |
|-----------|-----------|-----------|-----------|
| Sandstone | 31405 | 929 | 645 |
| Claystone | 84 | 6762 | 0 |

Alle: 39825
Riktig: 38167
Feil: 1658
Feilrate: 0,04163214
Klassifisering: 0,95836786

Antall sensorer man bør benytte på Logg 2 med Naive Bayes Classifier

Shallow Resistivity

| | Sandstone | Claystone | Limestone |
|-----------|-----------|-----------|-----------|
| Sandstone | 26413 | 4015 | 228 |
| Claystone | 62 | 6770 | 0 |

Alle: 37488
Riktig: 33183
Feil: 4305
Feilrate: 0,11483675
Klassifisering: 0,88516325

Shallow Resistivity, Medium Resistivity / Mud Resistivity from HALS Measurements

| | Sandstone | Claystone | Limestone |
|-----------|-----------|-----------|-----------|
| Sandstone | 29030 | 1336 | 293 |
| Claystone | 76 | 6757 | 0 |

Alle: 37492
Riktig: 35787
Feil: 1705
Feilrate: 0,04547637
Klassifisering: 0,95452363

Shallow Resistivity, Medium Resistivity / Mud Resistivity from HALS Measurements, Deep Resistivity

| | Sandstone | Claystone | Limestone |
|-----------|-----------|-----------|-----------|
| Sandstone | 29110 | 1111 | 441 |
| Claystone | 76 | 6757 | 0 |

Alle: 37495
Riktig: 35867
Feil: 1628
Feilrate: 0,04341912
Klassifisering: 0,95658088

Shallow Resistivity, Medium Resistivity / Mud Resistivity from HALS Measurements, Deep Resistivity, Thermal Neutron Porosity

| | Sandstone | Claystone | Limestone |
|-----------|-----------|-----------|-----------|
| Sandstone | 31600 | 978 | 400 |
| Claystone | 76 | 6757 | 0 |

Alle: 39811
Riktig: 38357
Feil: 1454
Feilrate: 0,03652257
Klassifisering: 0,96347743

Shallow Resistivity, Medium Resistivity / Mud Resistivity from HALS Measurements, Deep Resistivity, Thermal Neutron Porosity, High Resolution Bulk Density

| | Sandstone | Claystone | Limestone |
|-----------|-----------|-----------|-----------|
| Sandstone | 31626 | 980 | 372 |
| Claystone | 76 | 6757 | 0 |

Alle: 39811
Riktig: 38383
Feil: 1428
Feilrate: 0,03586948
Klassifisering: 0,96413052

Shallow Resistivity, Medium Resistivity / Mud Resistivity from HALS Measurements, Deep Resistivity, Thermal Neutron Porosity, High Resolution Bulk Density, High Resolution Gamma Ray

| | Sandstone | Claystone | Limestone |
|-----------|-----------|-----------|-----------|
| Sandstone | 31775 | 914 | 289 |
| Claystone | 83 | 6763 | 0 |

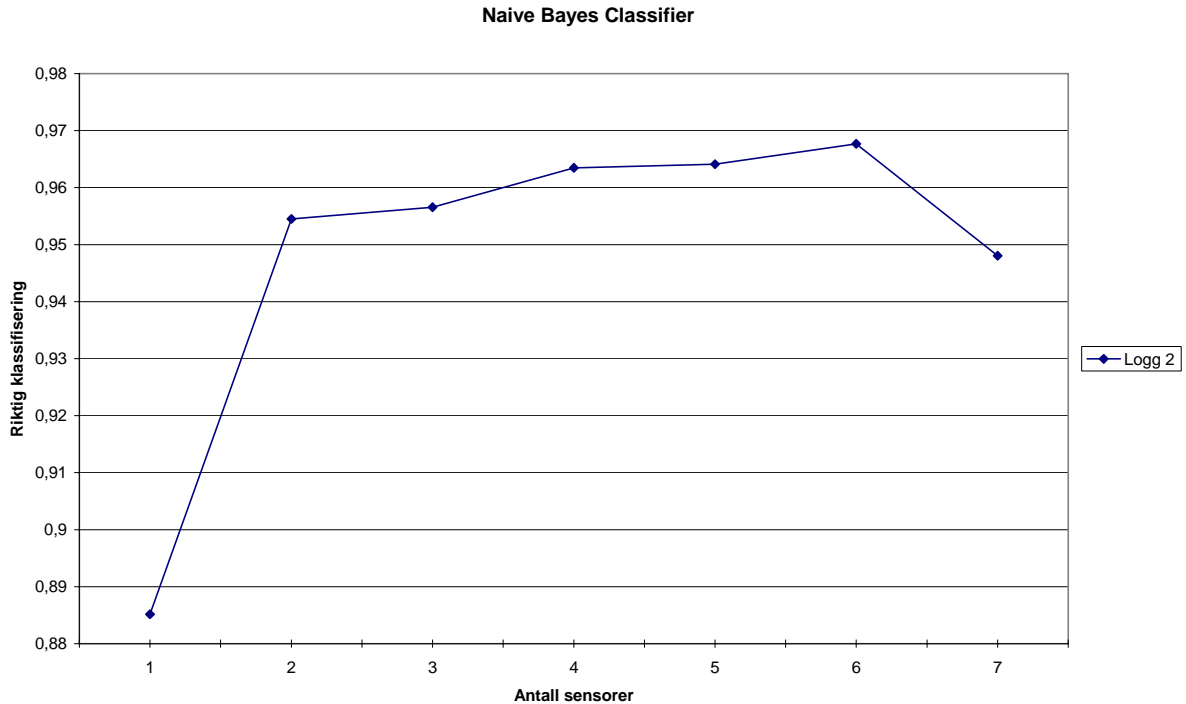
Alle: 39824
Riktig: 38538
Feil: 1286
Feilrate: 0,03229209
Klassifisering: 0,96770791

Shallow Resistivity, Medium Resistivity / Mud Resistivity from HALS Measurements, Deep Resistivity, Thermal Neutron Porosity, High Resolution Bulk Density, High Resolution Gamma Ray, Medium Shallow Resistivity (lik Naive Bayes Classifier over)

| | Sandstone | Claystone | Limestone |
|-----------|-----------|-----------|-----------|
| Sandstone | 30993 | 1612 | 375 |
| Claystone | 82 | 6764 | 0 |

Alle: 39826
Riktig: 37757
Feil: 2069
Feilrate: 0,05195099
Klassifisering: 0,94804901

Grafisk fremstilling av resultatene



Antall sensorer man bør benytte på Logg 2 med Hidden Markov Model

Shallow Resistivity

| | Sandstone | Claystone | Limestone |
|-----------|-----------|-----------|-----------|
| Sandstone | 26708 | 3717 | 231 |
| Claystone | 52 | 6780 | 0 |

Alle: 37488
Riktig: 33488
Feil: 4000
Feilrate: 0,10670081
Klassifisering: 0,89329919

Shallow Resistivity, Medium Resistivity / Mud Resistivity from HALS Measurements

| | Sandstone | Claystone | Limestone |
|-----------|-----------|-----------|-----------|
| Sandstone | 29065 | 1289 | 305 |
| Claystone | 75 | 6758 | 0 |

Alle: 37492
Riktig: 35823
Feil: 1669
Feilrate: 0,04451616
Klassifisering: 0,95548384

Shallow Resistivity, Medium Resistivity / Mud Resistivity from HALS Measurements, Deep Resistivity

| | Sandstone | Claystone | Limestone |
|-----------|-----------|-----------|-----------|
| Sandstone | 29181 | 1051 | 430 |
| Claystone | 74 | 6759 | 0 |

Alle: 37495
Riktig: 35940
Feil: 1555
Feilrate: 0,0414722
Klassifisering: 0,9585278

Shallow Resistivity, Medium Resistivity / Mud Resistivity from HALS Measurements, Deep Resistivity, Thermal Neutron Porosity

| | Sandstone | Claystone | Limestone |
|-----------|-----------|-----------|-----------|
| Sandstone | 31641 | 936 | 401 |
| Claystone | 74 | 6759 | 0 |

Alle: 39811
Riktig: 38400
Feil: 1411
Feilrate: 0,03544247
Klassifisering: 0,96455753

Shallow Resistivity, Medium Resistivity / Mud Resistivity from HALS Measurements, Deep Resistivity, Thermal Neutron Porosity, High Resolution Bulk Density

| | Sandstone | Claystone | Limestone |
|-----------|-----------|-----------|-----------|
| Sandstone | 31668 | 936 | 374 |
| Claystone | 75 | 6758 | 0 |

Alle: 39811
Riktig: 38426
Feil: 1385
Feilrate: 0,03478938
Klassifisering: 0,96521062

Shallow Resistivity, Medium Resistivity / Mud Resistivity from HALS Measurements, Deep Resistivity, Thermal Neutron Porosity, High Resolution Bulk Density, High Resolution Gamma Ray

| | Sandstone | Claystone | Limestone |
|-----------|-----------|-----------|-----------|
| Sandstone | 31834 | 897 | 247 |
| Claystone | 75 | 6771 | 0 |

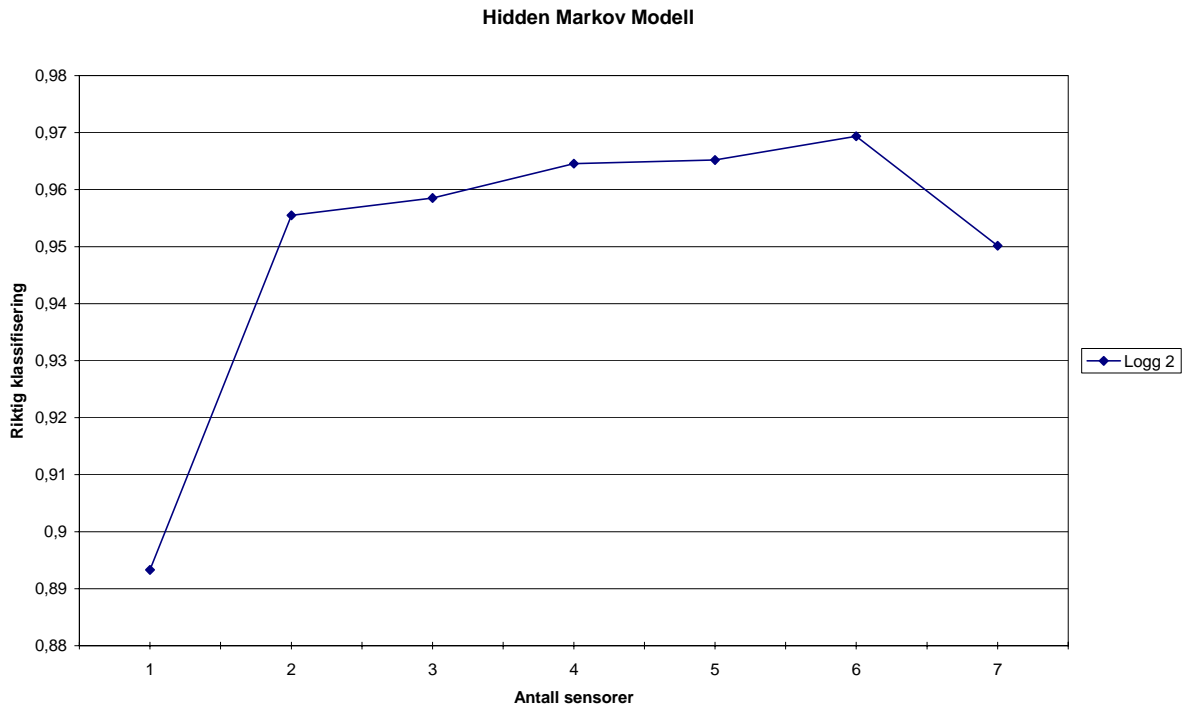
Alle: 39824
Riktig: 38605
Feil: 1219
Feilrate: 0,03060968
Klassifisering: 0,96939032

Shallow Resistivity, Medium Resistivity / Mud Resistivity from HALS Measurements, Deep Resistivity, Thermal Neutron Porosity, High Resolution Bulk Density, High Resolution Gamma Ray, Medium Shallow Resistivity (lik Hidden Markov Model over)

| | Sandstone | Claystone | Limestone |
|-----------|-----------|-----------|-----------|
| Sandstone | 31070 | 1548 | 362 |
| Claystone | 75 | 6771 | 0 |

Alle: 39826
Riktig: 37841
Feil: 1985
Feilrate: 0,04984181
Klassifisering: 0,95015819

Grafisk fremstilling av resultatene



Naive Bayes Classifier når treningsgrunnet er beregnet fra andre logger

| | Sandstone | Claystone |
|-----------|-----------|-----------|
| Sandstone | 0 | 32980 |
| Claystone | 0 | 6846 |

Alle: 39826
Riktig: 6846
Feil: 32980
Feilrate: 0,82810224
Klassifisering: 0,17189776

Hidden Markov Model når treningsgrunnet er beregnet fra andre logger

| | Sandstone | Claystone |
|-----------|-----------|-----------|
| Sandstone | 0 | 32980 |
| Claystone | 0 | 6846 |

Alle: 39826
Riktig: 6846
Feil: 32980
Feilrate: 0,82810224
Klassifisering: 0,17189776

Hidden Markov Model med snittovergangsmatrise når treningsgrunnet er beregnet fra andre logger

| | Sandstone | Claystone |
|-----------|-----------|-----------|
| Sandstone | 0 | 32980 |
| Claystone | | 6846 |

Alle: 39826
Riktig: 6846
Feil: 32980
Feilrate: 0,82810224
Klassifisering: 0,17189776

Vedlegg C – Forvirringsmatriser for Logg 3

Enkelt sensorer

Deep Resistivity

| | Sandstone | Limestone | Claystone |
|-----------|-----------|-----------|-----------|
| Sandstone | 801 | 240 | 1932 |
| Limestone | 5 | 243 | 141 |
| Claystone | 23 | 1152 | 11316 |

Alle: 15853
Riktig: 12360
Feil: 3493
Feilrate: 0,22033684
Klassifisering: 0,77966316

High Resolution Bulk Density

| | Sandstone | Limestone | Claystone |
|-----------|-----------|-----------|-----------|
| Sandstone | 921 | 580 | 1370 |
| Limestone | 25 | 124 | 166 |
| Claystone | 3 | 21 | 266 |

Alle: 3476
Riktig: 1311
Feil: 2165
Feilrate: 0,62284235
Klassifisering: 0,37715765

High Resolution Gamma Ray

| | Sandstone | Limestone | Claystone |
|-----------|-----------|-----------|-----------|
| Sandstone | 1154 | 1078 | 769 |
| Limestone | 185 | 221 | 21 |
| Claystone | 4036 | 3525 | 4982 |

Alle: 15971
Riktig: 6357
Feil: 9614
Feilrate: 0,60196606
Klassifisering: 0,39803394

Medium Resistivity / Mud Resistivity from HALS Measurements

| | Sandstone | Limestone | Claystone |
|-----------|-----------|-----------|-----------|
| Sandstone | 121 | 768 | 2115 |
| Limestone | 16 | 10 | 363 |
| Claystone | 0 | 53 | 12476 |

Alle: 15922
Riktig: 12607
Feil: 3315
Feilrate: 0,20820249
Klassifisering: 0,79179751

Thermal Neutron Porosity

| | Sandstone | Limestone | Claystone |
|-----------|-----------|-----------|-----------|
| Sandstone | 1094 | 1397 | 332 |
| Limestone | 101 | 138 | 57 |
| Claystone | 36 | 20 | 72 |

Alle: 3247
Riktig: 1304
Feil: 1943
Feilrate: 0,59839852
Klassifisering: 0,40160148

Naive Bayes Classifier

| | Sandstone | Limestone | Claystone | Shale |
|-----------|-----------|-----------|-----------|-------|
| Sandstone | 831 | 338 | 1831 | 4 |
| Limestone | 25 | 196 | 229 | 0 |
| Claystone | 28 | 289 | 12226 | 0 |

Alle: 15997
Riktig: 13253
Feil: 2744
Feilrate: 0,17153216
Klassifisering: 0,82846784

Hidden Markov Model

| | Sandstone | Limestone | Claystone | Shale |
|-----------|-----------|-----------|-----------|-------|
| Sandstone | 834 | 337 | 1830 | 3 |
| Limestone | 16 | 193 | 241 | 0 |
| Claystone | 14 | 260 | 12269 | 0 |

Alle: 15997
Riktig: 13296
Feil: 2701
Feilrate: 0,16884416
Klassifisering: 0,83115584

Hidden Markov Model med snittovergangsmatrise

| | Sandstone | Limestone | Claystone | Siltstone |
|-----------|-----------|-----------|-----------|-----------|
| Sandstone | 833 | 337 | 1830 | 4 |
| Limestone | 13 | 194 | 243 | 0 |
| Claystone | 14 | 260 | 12269 | 0 |

Alle: 15997
Riktig: 13296
Feil: 2701
Feilrate: 0,16884416
Klassifisering: 0,83115584

Naive Bayes Classifier når 1 % av uteliggerne er fjernet

| | Sandstone | Limestone | Claystone | Shale |
|-----------|-----------|-----------|-----------|-------|
| Sandstone | 1045 | 264 | 1695 | 0 |
| Limestone | 19 | 302 | 125 | 4 |
| Claystone | 82 | 617 | 11844 | 0 |

Alle: 15997
Riktig: 13191
Feil: 2806
Feilrate: 0,17540789
Klassifisering: 0,82459211

Hidden Markov Model når 1 % av uteliggerne er fjernet

| | Sandstone | Limestone | Claystone | Shale |
|-----------|-----------|-----------|-----------|-------|
| Sandstone | 1029 | 276 | 1699 | 0 |
| Limestone | 2 | 301 | 143 | 4 |
| Claystone | 55 | 555 | 11933 | 0 |

Alle: 15997
Riktig: 13263
Feil: 2734
Feilrate: 0,17090705
Klassifisering: 0,82909295

Naive Bayes Classifier når 1 % av uteliggerne er fjernet

| | Sandstone | Limestone | Claystone | Shale |
|-----------|-----------|-----------|-----------|-------|
| Sandstone | 838 | 350 | 1810 | 6 |
| Limestone | 26 | 204 | 220 | 0 |
| Claystone | 30 | 296 | 12217 | 0 |

Alle: 15997
Riktig: 13259
Feil: 2738
Feilrate: 0,17115709
Klassifisering: 0,82884291

Hidden Markov Model når 1 % av uteliggerne er fjernet

| | Sandstone | Limestone | Claystone | Shale |
|-----------|-----------|-----------|-----------|-------|
| Sandstone | 843 | 345 | 1810 | 6 |
| Limestone | 16 | 208 | 226 | 0 |
| Claystone | 17 | 279 | 12247 | 0 |

Alle: 15997
Riktig: 13298
Feil: 2699
Feilrate: 0,16871913
Klassifisering: 0,83128087

Antall sensorer man bør benytte på Logg 3 med Naive Bayes Classifier

Medium Resistivity / Mud Resistivity from HALS Measurements

| | Sandstone | Limestone | Claystone | Shale |
|-----------|-----------|-----------|-----------|-------|
| Sandstone | 116 | 768 | 2115 | 5 |
| Limestone | 16 | 10 | 363 | 0 |
| Claystone | 0 | 53 | 12476 | 0 |

Alle: 15922
Riktig: 12602
Feil: 3320
Feilrate: 0,20851652
Klassifisering: 0,79148348

Medium Resistivity / Mud Resistivity from HALS Measurements, Deep Resistivity

| | Sandstone | Limestone | Claystone | Shale |
|-----------|-----------|-----------|-----------|-------|
| Sandstone | 818 | 330 | 1849 | 7 |
| Limestone | 13 | 111 | 267 | 0 |
| Claystone | 10 | 258 | 12262 | 0 |

Alle: 15925
Riktig: 13191
Feil: 2734
Feilrate: 0,17167975
Klassifisering: 0,82832025

Medium Resistivity / Mud Resistivity from HALS Measurements, Deep Resistivity, Thermal Neutron Porosity

| | Sandstone | Limestone | Claystone | Shale |
|-----------|-----------|-----------|-----------|-------|
| Sandstone | 826 | 326 | 1847 | 5 |
| Limestone | 13 | 143 | 265 | 0 |
| Claystone | 13 | 255 | 12262 | 0 |
| | | | | |

Alle: 15955
Riktig: 13231
Feil: 2724
Feilrate: 0,17073018
Klassifisering: 0,82926982

Medium Resistivity / Mud Resistivity from HALS Measurements, Deep Resistivity, Thermal Neutron Porosity, High Resolution Bulk Density

| | Sandstone | Limestone | Claystone | Shale |
|-----------|-----------|-----------|-----------|-------|
| Sandstone | 828 | 338 | 1834 | 4 |
| Limestone | 14 | 172 | 235 | 0 |
| Claystone | 12 | 255 | 12263 | 0 |

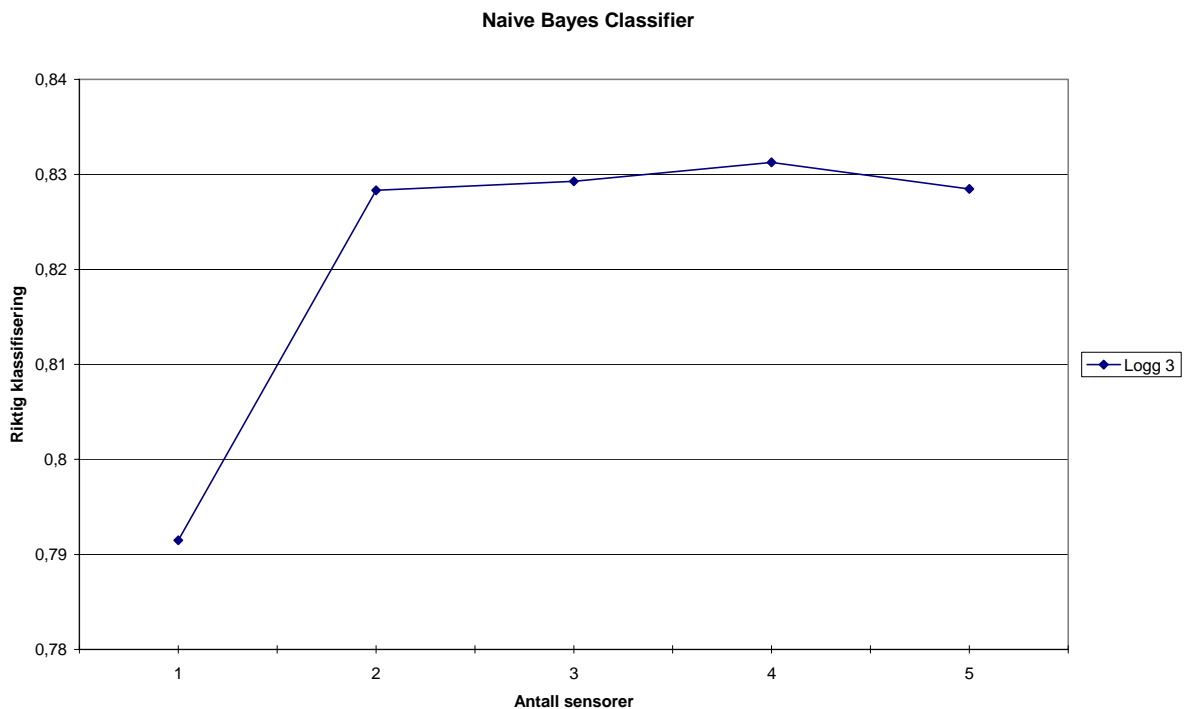
Alle: 15955
Riktig: 13263
Feil: 2692
Feilrate: 0,16872454
Klassifisering: 0,83127546

Medium Resistivity / Mud Resistivity from HALS Measurements, Deep Resistivity, Thermal Neutron Porosity, High Resolution Bulk Density, High Resolution Gamma Ray (lik Naive Bayes Classifier over)

| | Sandstone | Limestone | Claystone | Shale |
|-----------|-----------|-----------|-----------|-------|
| Sandstone | 831 | 338 | 1831 | 4 |
| Limestone | 25 | 196 | 229 | 0 |
| Claystone | 28 | 289 | 12226 | 0 |

Alle: 15997
Riktig: 13253
Feil: 2744
Feilrate: 0,17153216
Klassifisering: 0,82846784

Grafisk fremstilling av resultatene



Antall sensorer man bør benytte på Logg 3 med Hidden Markov Model

Medium Resistivity / Mud Resistivity from HALS Measurements

| | Sandstone | Limestone | Claystone | Shale |
|-----------|-----------|-----------|-----------|-------|
| Sandstone | 82 | 822 | 2097 | 3 |
| Limestone | 20 | 15 | 354 | 0 |
| Claystone | 0 | 51 | 12478 | 0 |

Alle: 15922
Riktig: 12575
Feil: 3347
Feilrate: 0,21021228
Klassifisering: 0,78978772

Medium Resistivity / Mud Resistivity from HALS Measurements, Deep Resistivity

| | Sandstone | Limestone | Claystone | Shale |
|-----------|-----------|-----------|-----------|-------|
| Sandstone | 829 | 321 | 1849 | 5 |
| Limestone | 18 | 111 | 262 | 0 |
| Claystone | 11 | 261 | 12258 | 0 |

Alle: 15925
Riktig: 13198
Feil: 2727
Feilrate: 0,17124019
Klassifisering: 0,82875981

Medium Resistivity / Mud Resistivity from HALS Measurements, Deep Resistivity, Thermal Neutron Porosity

| | Sandstone | Limestone | Claystone | Shale |
|-----------|-----------|-----------|-----------|-------|
| Sandstone | 828 | 325 | 1847 | 4 |
| Limestone | 18 | 161 | 242 | 0 |
| Claystone | 13 | 259 | 12258 | 0 |

Alle: 15955
Riktig: 13247
Feil: 2708
Feilrate: 0,16972736
Klassifisering: 0,83027264

Medium Resistivity / Mud Resistivity from HALS Measurements, Deep Resistivity, Thermal Neutron Porosity, High Resolution Bulk Density

| | Sandstone | Limestone | Claystone | Shale |
|-----------|-----------|-----------|-----------|-------|
| Sandstone | 832 | 334 | 1835 | 3 |
| Limestone | 19 | 154 | 248 | 0 |
| Claystone | 12 | 257 | 12261 | 0 |

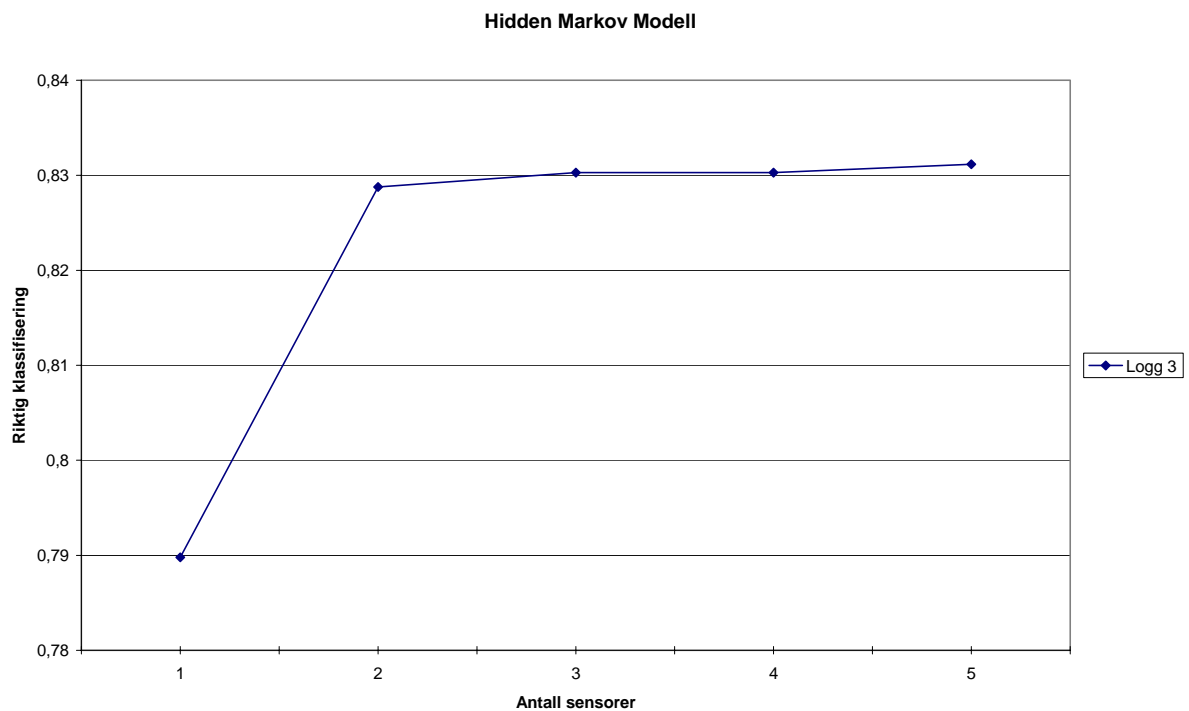
Alle: 15955
Riktig: 13247
Feil: 2708
Feilrate: 0,16972736
Klassifisering: 0,83027264

Medium Resistivity / Mud Resistivity from HALS Measurements, Deep Resistivity, Thermal Neutron Porosity, High Resolution Bulk Density, High Resolution Gamma Ray (lik Hidden Markov Model over)

| | Sandstone | Limestone | Claystone | Shale |
|-----------|-----------|-----------|-----------|-------|
| Sandstone | 834 | 337 | 1830 | 3 |
| Limestone | 16 | 193 | 241 | 0 |
| Claystone | 14 | 260 | 12269 | 0 |

Alle: 15997
Riktig: 13296
Feil: 2701
Feilrate: 0,16884416
Klassifisering: 0,83115584

Grafisk fremstilling av resultatene



Naive Bayes Classifier når treningsgrunnet er beregnet fra andre logger

| | Sandstone | Limestone | Claystone | Anhydrite | Siltstone | Dolomite |
|-----------|-----------|-----------|-----------|-----------|-----------|----------|
| Sandstone | 477 | 1175 | 1046 | 13 | 1 | 292 |
| Limestone | 57 | 186 | 170 | 13 | 17 | 7 |
| Claystone | 60 | 642 | 11837 | 0 | 0 | 4 |

Alle: 15997
Riktig: 12500
Feil: 3497
Feilrate: 0,21860349
Klassifisering: 0,78139651

Hidden Markov Model når treningsgrunnet er beregnet fra andre logger

| | Sandstone | Limestone | Claystone | Anhydrite | Siltstone | Dolomite |
|-----------|-----------|-----------|-----------|-----------|-----------|----------|
| Sandstone | 490 | 1149 | 1062 | 11 | 0 | 292 |
| Limestone | 59 | 165 | 186 | 15 | 17 | 8 |
| Claystone | 81 | 564 | 11895 | 0 | 0 | 3 |

Alle: 15997
Riktig: 12550
Feil: 3447
Feilrate: 0,2154779
Klassifisering: 0,7845221

Hidden Markov Model med snittovergangsmatrise når treningsgrunnet er beregnet fra andre logger

| | Sandstone | Limestone | Claystone | Anhydrite | Siltstone | Dolomite |
|-----------|-----------|-----------|-----------|-----------|-----------|----------|
| Sandstone | 490 | 1149 | 1062 | 11 | 0 | 292 |
| Limestone | 61 | 167 | 184 | 13 | 17 | 8 |
| Claystone | 78 | 564 | 11898 | 0 | 0 | 3 |

Alle: 15997
Riktig: 12555
Feil: 3442
Feilrate: 0,21516534
Klassifisering: 0,78483466

Vedlegg D – Forvirringsmatriser for Logg 4

Enkelt sensorer

Delta-T / Sonic

| | Sandstone | Limestone | Shale | Anhydrite |
|-----------|-----------|-----------|-------|-----------|
| Sandstone | 215 | 0 | 0 | 3 |
| Limestone | 0 | 423 | 0 | 29 |
| Shale | 0 | 0 | 1877 | 0 |
| Anhydrite | 0 | 5 | 0 | 208 |

Alle: 2760
Riktig: 2723
Feil: 37
Feilrate: 0,0134058
Klassifisering: 0,9865942

High Resolution Bulk Density

| | Sandstone | Limestone | Shale | Anhydrite |
|-----------|-----------|-----------|-------|-----------|
| Sandstone | 203 | 0 | 0 | 15 |
| Limestone | 0 | 441 | 0 | 11 |
| Shale | 0 | 0 | 0 | 0 |
| Anhydrite | 1 | 0 | 0 | 162 |

Alle: 833
Riktig: 806
Feil: 27
Feilrate: 0,03241297
Klassifisering: 0,96758703

High Resolution Gamma Ray

| | Sandstone | Limestone | Shale | Anhydrite |
|-----------|-----------|-----------|-------|-----------|
| Sandstone | 213 | 5 | 0 | 0 |
| Limestone | 21 | 284 | 0 | 147 |
| Shale | 0 | 0 | 1877 | 0 |
| Anhydrite | 1 | 111 | 0 | 101 |

Alle: 2760
Riktig: 2475
Feil: 285
Feilrate: 0,10326087
Klassifisering: 0,89673913

Medium Resistivity / Mud Resistivity from HALS Measurements

| | Sandstone | Limestone | Shale | Anhydrite |
|-----------|-----------|-----------|-------|-----------|
| Sandstone | 212 | 0 | 6 | 0 |
| Limestone | 0 | 449 | 0 | 3 |
| Shale | 21 | 4 | 1852 | 0 |
| Anhydrite | 0 | 0 | 0 | 213 |

Alle: 2760
Riktig: 2726
Feil: 34
Feilrate: 0,01231884
Klassifisering: 0,98768116

Thermal Neutron Porosity

| | Sandstone | Limestone | Shale | Anhydrite |
|-----------|-----------|-----------|-------|-----------|
| Sandstone | 214 | 4 | 0 | 0 |
| Limestone | 1 | 450 | 0 | 1 |
| Shale | 0 | 0 | 0 | 0 |
| Anhydrite | 0 | 3 | 0 | 174 |

Alle: 847
Riktig: 838
Feil: 9
Feilrate: 0,01062574
Klassifisering: 0,98937426

Naive Bayes Classifier

| | Sandstone | Limestone | Shale | Anhydrite | Claystone |
|-----------|-----------|-----------|-------|-----------|-----------|
| Sandstone | 214 | 0 | 0 | 0 | 3 |
| Limestone | 0 | 446 | 0 | 3 | 2 |
| Shale | 0 | 0 | 1875 | 0 | 0 |
| Anhydrite | 0 | 0 | 0 | 211 | 4 |

Alle: 2758
Riktig: 2746
Feil: 12
Feilrate: 0,00435098
Klassifisering: 0,99564902

Hidden Markov Modell

| | Sandstone | Limestone | Shale | Anhydrite | Claystone |
|-----------|-----------|-----------|-------|-----------|-----------|
| Sandstone | 214 | 0 | 0 | 0 | 3 |
| Limestone | 0 | 444 | 0 | 5 | 2 |
| Shale | 0 | 0 | 1875 | 0 | 0 |
| Anhydrite | 0 | 0 | 0 | 211 | 4 |

Alle: 2758
Riktig: 2744
Feil: 14
Feilrate: 0,00507614
Klassifisering: 0,99492386

Hidden Markov Model med snittovergangsmatrise

| | Sandstone | Limestone | Shale | Anhydrite | Siltstone |
|-----------|-----------|-----------|-------|-----------|-----------|
| Sandstone | 214 | 0 | 0 | 0 | 3 |
| Limestone | 0 | 444 | 0 | 5 | 2 |
| Shale | 0 | 0 | 1875 | 0 | 0 |
| Anhydrite | 0 | 0 | 0 | 211 | 4 |

Alle: 2758
Riktig: 2744
Feil: 14
Feilrate: 0,00507614
Klassifisering: 0,99492386

Naive Bayes Classifier når 1 % av uteliggerne er fjernet

| | Sandstone | Limestone | Shale | Anhydrite | Claystone |
|-----------|-----------|-----------|-------|-----------|-----------|
| Sandstone | 215 | 0 | 0 | 0 | 2 |
| Limestone | 0 | 449 | 0 | 0 | 2 |
| Shale | 0 | 0 | 1873 | 0 | 2 |
| Anhydrite | 0 | 2 | 0 | 210 | 3 |

Alle: 2758
Riktig: 2747
Feil: 11
Feilrate: 0,0039884
Klassifisering: 0,9960116

Hidden Markov Model når 1 % av uteliggerne er fjernet

| | Sandstone | Limestone | Shale | Anhydrite | Claystone |
|-----------|-----------|-----------|-------|-----------|-----------|
| Sandstone | 215 | 0 | 0 | 0 | 2 |
| Limestone | 0 | 449 | 0 | 0 | 2 |
| Shale | 0 | 0 | 1875 | 0 | 0 |
| Anhydrite | 0 | 0 | 0 | 212 | 3 |

Alle: 2758
Riktig: 2751
Feil: 7
Feilrate: 0,00253807
Klassifisering: 0,99746193

Naive Bayes Classifier når 1 % av uteliggerne er fjernet

| | Sandstone | Limestone | Shale | Anhydrite | Claystone |
|-----------|-----------|-----------|-------|-----------|-----------|
| Sandstone | 214 | 0 | 0 | 0 | 3 |
| Limestone | 0 | 446 | 0 | 3 | 2 |
| Shale | 0 | 0 | 1875 | 0 | 0 |
| Anhydrite | 0 | 0 | 0 | 211 | 4 |

Alle: 2758
Riktig: 2746
Feil: 12
Feilrate: 0,00435098
Klassifisering: 0,99564902

Hidden Markov Model når 1 % av uteliggerne er fjernet

| | Sandstone | Limestone | Shale | Anhydrite | Claystone |
|-----------|-----------|-----------|-------|-----------|-----------|
| Sandstone | 214 | 0 | 0 | 0 | 3 |
| Limestone | 0 | 444 | 0 | 5 | 2 |
| Shale | 0 | 0 | 1875 | 0 | 0 |
| Anhydrite | 0 | 0 | 0 | 211 | 4 |

Alle: 2758
Riktig: 2744
Feil: 14
Feilrate: 0,00507614
Klassifisering: 0,99492386

Antall sensorer man bør benytte på Logg 4 med Naive Bayes Classifier

Delta-T / Sonic

| | Sandstone | Limestone | Shale | Anhydrite | Claystone |
|-----------|-----------|-----------|-------|-----------|-----------|
| Sandstone | 214 | 0 | 0 | 3 | 0 |
| Limestone | 0 | 423 | 0 | 28 | 0 |
| Shale | 0 | 0 | 1866 | 0 | 9 |
| Anhydrite | 0 | 5 | 0 | 208 | 0 |

Alle: 2756
Riktig: 2711
Feil: 45
Feilrate: 0,01632801
Klassifisering: 0,98367199

Delta-T / Sonic, Resistivity Shallow

| | Sandstone | Limestone | Shale | Anhydrite | Claystone |
|-----------|-----------|-----------|-------|-----------|-----------|
| Sandstone | 213 | 0 | 0 | 4 | 0 |
| Limestone | 0 | 423 | 1 | 27 | 0 |
| Shale | 0 | 0 | 1872 | 0 | 3 |
| Anhydrite | 1 | 5 | 0 | 208 | 0 |

Alle: 2757
Riktig: 2716
Feil: 41
Feilrate: 0,01487124
Klassifisering: 0,98512876

Delta-T / Sonic, Resistivity Shallow, Medium Resistivity / Mud Resistivity from HALS Measurements

| | Sandstone | Limestone | Shale | Anhydrite | Claystone |
|-----------|-----------|-----------|-------|-----------|-----------|
| Sandstone | 215 | 0 | 0 | 2 | 0 |
| Limestone | 0 | 447 | 0 | 4 | 0 |
| Shale | 0 | 0 | 1875 | 0 | 0 |
| Anhydrite | 0 | 0 | 0 | 215 | 0 |

Alle: 2758
Riktig: 2752
Feil: 6
Feilrate: 0,00217549
Klassifisering: 0,99782451

Delta-T / Sonic, Resistivity Shallow, Medium Resistivity / Mud Resistivity from HALS Measurements, Thermal Neutron Porosity

| | Sandstone | Limestone | Shale | Anhydrite | Claystone |
|-----------|-----------|-----------|-------|-----------|-----------|
| Sandstone | 215 | 0 | 0 | 0 | 2 |
| Limestone | 0 | 449 | 0 | 0 | 2 |
| Shale | 0 | 0 | 1875 | 0 | 0 |
| Anhydrite | 0 | 0 | 0 | 212 | 3 |

Alle: 2758
Riktig: 2751
Feil: 7
Feilrate: 0,00253807
Klassifisering: 0,99746193

Delta-T / Sonic, Resistivity Shallow, Medium Resistivity / Mud Resistivity from HALS Measurements, Thermal Neutron Porosity, High Resolution Bulk Density

| | Sandstone | Limestone | Shale | Anhydrite | Claystone |
|-----------|-----------|-----------|-------|-----------|-----------|
| Sandstone | 215 | 0 | 0 | 0 | 2 |
| Limestone | 0 | 446 | 0 | 5 | 0 |
| Shale | 0 | 0 | 1875 | 0 | 0 |
| Anhydrite | 0 | 0 | 0 | 213 | 2 |

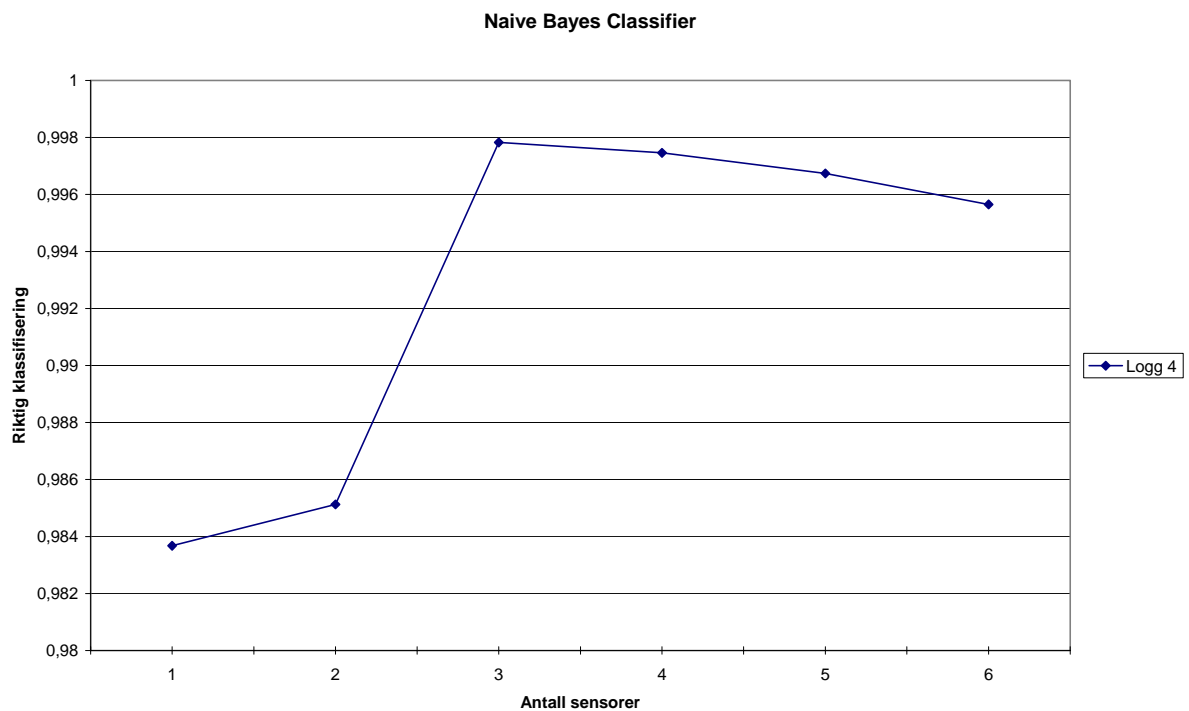
Alle: 2758
Riktig: 2749
Feil: 9
Feilrate: 0,00326323
Klassifisering: 0,99673677

Delta-T / Sonic, Resistivity Shallow, Medium Resistivity / Mud Resistivity from HALS Measurements, Thermal Neutron Porosity, High Resolution Bulk Density, High Resolution Gamma Ray (lik Naive Bayes Classifier over)

| | Sandstone | Limestone | Shale | Anhydrite | Claystone |
|-----------|-----------|-----------|-------|-----------|-----------|
| Sandstone | 214 | 0 | 0 | 0 | 3 |
| Limestone | 0 | 446 | 0 | 3 | 2 |
| Shale | 0 | 0 | 1875 | 0 | 0 |
| Anhydrite | 0 | 0 | 0 | 211 | 4 |

Alle: 2758
Riktig: 2746
Feil: 12
Feilrate: 0,00435098
Klassifisering: 0,99564902

Grafisk fremstilling av resultatene



Antall sensorer man bør benytte på Logg 4 med Hidden Markov Model

Delta-T / Sonic

| | Sandstone | Limestone | Shale | Anhydrite | Claystone |
|-----------|-----------|-----------|-------|-----------|-----------|
| Sandstone | 216 | 0 | 0 | 1 | 0 |
| Limestone | 0 | 442 | 0 | 9 | 0 |
| Shale | 0 | 0 | 1871 | 0 | 4 |
| Anhydrite | 0 | 0 | 0 | 213 | 0 |

Alle: 2756
Riktig: 2742
Feil: 14
Feilrate: 0,00507983
Klassifisering: 0,99492017

Delta-T / Sonic, Resistivity Shallow

| | Sandstone | Limestone | Shale | Anhydrite |
|-----------|-----------|-----------|-------|-----------|
| Sandstone | 215 | 0 | 0 | 2 |
| Limestone | 0 | 442 | 1 | 8 |
| Shale | 0 | 0 | 1875 | 0 |
| Anhydrite | 1 | 0 | 0 | 213 |

Alle: 2757
Riktig: 2745
Feil: 12
Feilrate: 0,00435256
Klassifisering: 0,99564744

Delta-T / Sonic, Resistivity Shallow, Medium Resistivity / Mud Resistivity from HALS Measurements

| | Sandstone | Limestone | Shale | Anhydrite |
|-----------|-----------|-----------|-------|-----------|
| Sandstone | 215 | 0 | 0 | 2 |
| Limestone | 0 | 447 | 0 | 4 |
| Shale | 0 | 0 | 1875 | 0 |
| Anhydrite | 0 | 0 | 0 | 215 |

Alle: 2758
Riktig: 2752
Feil: 6
Feilrate: 0,00217549
Klassifisering: 0,99782451

Delta-T / Sonic, Resistivity Shallow, Medium Resistivity / Mud Resistivity from HALS Measurements, Thermal Neutron Porosity

| | Sandstone | Limestone | Shale | Anhydrite | Claystone |
|-----------|-----------|-----------|-------|-----------|-----------|
| Sandstone | 215 | 0 | 0 | 0 | 2 |
| Limestone | 0 | 449 | 0 | 0 | 2 |
| Shale | 0 | 0 | 1875 | 0 | 0 |
| Anhydrite | 0 | 0 | 0 | 213 | 2 |

Alle: 2758
Riktig: 2752
Feil: 6
Feilrate: 0,00217549
Klassifisering: 0,99782451

Delta-T / Sonic, Resistivity Shallow, Medium Resistivity / Mud Resistivity from HALS Measurements, Thermal Neutron Porosity, High Resolution Bulk Density

| | Sandstone | Limestone | Shale | Anhydrite | Claystone |
|-----------|-----------|-----------|-------|-----------|-----------|
| Sandstone | 215 | 0 | 0 | 0 | 2 |
| Limestone | 0 | 446 | 0 | 5 | 0 |
| Shale | 0 | 0 | 1875 | 0 | 0 |
| Anhydrite | 0 | 0 | 0 | 213 | 2 |

Alle: 2758
Riktig: 2749
Feil: 9
Feilrate: 0,00326323
Klassifisering: 0,99673677

Delta-T / Sonic, Resistivity Shallow, Medium Resistivity / Mud Resistivity from HALS Measurements, Thermal Neutron Porosity, High Resolution Bulk Density, High Resolution Gamma Ray (lik Hidden Markov Model over)

| | Sandstone | Limestone | Shale | Anhydrite | Claystone |
|-----------|-----------|-----------|-------|-----------|-----------|
| Sandstone | 214 | 0 | 0 | 0 | 3 |
| Limestone | 0 | 444 | 0 | 5 | 2 |
| Shale | 0 | 0 | 1875 | 0 | 0 |
| Anhydrite | 0 | 0 | 0 | 211 | 4 |

Alle: 2758
Riktig: 2744
Feil: 14
Feilrate: 0,00507614
Klassifisering: 0,99492386

Grafisk fremstilling av resultatene

