



Norges miljø- og
biovitenskapelige
universitet

Masteroppgave 2022 30 stp
Handelshøyskolen

Bruk av metoder innen forklarbar maskinlæring for kundefracfallsprediksjon

Applying methods within explainable machine
learning for customer churn prediction

Abiramy Sivasubramaniam
Økonomi og administrasjon

Abstract

Artificial intelligence-based systems are constantly being used in decision-making processes. The concepts behind the systems are not always interpretable for humans. By using methods within explainable machine learning, it is possible to get a better understanding of these models.

This thesis tries to explore whether methods within explainable machine learning such as SHAP can be used to understand which factors that influence machine learning models within customer churn prediction. The focus is to understand the feature importance on these models at both local and global level. The results shows that the features which influence the model, do not necessarily have the same impact on the individual predictions.

Sammendrag

Kunstig intelligens baserte systemer blir stadig anvendt i beslutningstakingsprosesser. Det er ikke lett for mennesker å forstå de grunnleggende prinsippene bak hvordan disse systemene fungerer. Ved å anvende ulike metoder innen forklarbar maskinlæring er det mulig å få en bedre av forståelse av disse modellene.

I denne oppgaven undersøkes det hvorvidt metoder innen forklarbar maskinlæring som SHAP, kan anvendes for å forstå hvilke faktorer som påvirker maskinlæringsmodeller for kundefrafallsprediksjon i telekommunikasjonsbransjen. Det blir undersøkt hvilke variabler som påvirker modellene på både lokalt og globalt nivå. Analysen viser at variablene som påvirker selve modellen, nødvendigvis ikke har like stor påvirkning på de individuelle prediksjonene.

Innholdsfortegnelse

1. Introduksjon	6
2. Problemstilling.....	7
3. Teori	7
3.1. Kundefrafall	8
3.2. Kostnader knyttet til kundefrafall	8
3.3. Kundefrafall i telekommunikasjonsbransjen.....	9
3.4. Maskinlæring	9
3.4.1. Decision Tree	10
3.4.2. Random Forest.....	10
3.4.3. Support Vector Machine	11
3.4.4. Isolation Forest.....	11
3.4.5. Avveining mellom bias og varians	12
3.4.6. Evalueringskriterier.....	12
3.5. Forklarbarhet i maskinlæring.....	14
3.5.1. Hvorfor er det nødvendig med forklarbarhet i kundefrafallsprediksjon?	15
3.5.2. Metoder innen forklarbar maskinlæring.....	15
3.5.3. SHAP – Shapley Additive Explanation.....	17
3.5.4. LIME.....	18
4. Litteraturgjennomgang	18
4.1. Prediksjon av kundefrafall.....	18
4.2. Forklarbare maskinlæring innen kundefrafallsprediksjon.....	20
5. Metode.....	20
5.1. Data	20
5.2. Forbehandling av data	23
5.3. SHAP	27
6. Resultat.....	27
6.1. Optimale hyperparameter.....	27
6.2. Forvirringsmatrise.....	28

6.2.3. Evaluering.....	29
6.3. SHAP for global forståelse av modellen.....	30
6.3.1. Variablenes betydning for modellens prediksjoner.....	30
6.4. SHAP for lokal forståelse av modellen.....	33
6.4.1. Feil positiv (FP) – SHAP.....	33
6.4.2. Enkelttilfelle med kundefrafall på prediksjonsnivå.....	34
7. Diskusjon.....	34
8. Konklusjon.....	35
Referanser.....	37

1. Introduksjon

Telekommunikasjonsbransjen har vært gjennom store endringer i løpet av årene. Itarten av 2022 var det forventet at bransjen ville møte på nye muligheter og utfordringer som følge av regulatoriske, teknologiske- og konkurranserelaterte faktorer (Deloitte, 2022). Med økt konkurranse i telekommunikasjonsbransjen har det blitt enklere for kunder å bytte over til konkurrenter (Óskarsdóttir *et al.*, 2017). For å håndtere utfordringer knyttet til kundefrafall anvender selskaper ofte avansert teknologi innen maskinlæring.

For å predikere kundefrafall gjennom maskinlæringsmodeller er det viktig å kunne forstå hvilke faktorer som fører til kundefrafall (Berson *et al.*, 2000 sitert i Misra *et al.*, 2017). Dette er avgjørende for å kunne identifisere potensielle kundefrafallstilfeller raskere. Utfordringen med å anvende avanserte maskinlæringsmodeller er at prosessen bak modellene ikke er enkelt å forstå for mennesker (Rudin, 2018). Dette kan blant annet føre til at det blir vanskeligere å oppdage hvilke faktorer som fører til kundefrafall, som videre vil gjøre arbeidet med å bevare kunder utfordrende.

Ettersom avanserte metoder innen kunstig intelligens anvendes i ulike områder for beslutningstaking, innførte det Europeiske Parlament i 2018 nye krav til bruk av automatiserte systemer i beslutningsprosesser. Det ble bestemt at mennesker har rett til å få en forklaring på logikken som er involvert i automatiserte beslutningsprosesser. Som følge av dette har det de siste årene oppstått en større interesse for metoder for å forklare kunstig intelligensbaserte systemer. (Bodria *et al.*, 2021).

På bakgrunn av dette vil det i denne oppgaven anvendes ulike metoder innen forklarbar maskinlæring for å forstå predikeringutfallet til en kundefrafallsmodell. Det vil hovedsakelig fokuseres på å anvende metoder innen verktøyet *SHAP*.

2. Problemstilling

Med grunnlag i beskrivelsen i introduksjonen blir det formulert følgende problemstilling som vil danne grunnlaget for resten av oppgaven.

I hvilken grad kan metoder innen forklarbar maskinlæring bidra til økt forståelse av kundefrafallsprediksjon i telekommunikasjonsbransjen?

For å svare på problemstillingen blir det valgt å formulere følgende forskningsspørsmål:

- 1) Hvordan kan globale modell-agnostiske metoder innen forklarbar maskinlæring bidra til en bredere forståelse av hvilke faktorer som påvirker en maskinlæringsmodell innen kundefrafallsprediksjon?*
- 2) Hvordan kan lokale modell-agnostiske metoder innen forklarbar maskinlæring bidra til en bredere forståelse av hvilke faktorer som påvirker individuelle prediksjoner?*

Forskningsspørsmålene vil gjøre det enklere å kunne tilnærme problemstillingen, og anvende relevante metoder. Oppgaven starter med å presentere relevant teori, forklaring av metoder innen maskinlæring og forklarbar maskinlæring. Videre vil metode, analyse og resultater presenteres. Deretter diskuteres resultatene før problemstillingen besvares. Tilslutt kommer konklusjon og forslag til videre arbeid.

3. Teori

Denne seksjonen gir en oversikt over teoretisk bakgrunn som er relevant for denne oppgaven. Først vil teori innen kundefrafall, maskinlæring og forklarbar maskinlæring bli presentert. Deretter vil teori relatert til metoder som anvendes i denne oppgaven bli presentert.

3.1. Kundefrafall

Kundefrafall innebærer at et selskap mister en verdifull kunde som følge av at kunden har gått over til en konkurrent (Richeldi *et al.*, 2002; Huang 2012). Ifølge Hadden *et al.* (2007) kan kundefrafall deles inn i to ulike kategorier; frivillig- og ufrivillig kundefrafall. Frivillig kundefrafall oppstår når kunden bevisst eller tilfeldig tar en beslutning om å avslutte sitt kundeforhold hos selskapet. En bevisst avgjørelse i denne sammenhengen kan for eksempel være at kunden har funnet et bedre alternativ hos en konkurrent som kan være rimeligere, anvender nyere teknologi eller av bedre kvalitet. Kunder som bevisst avgjør å forlate selskapet er vanskeligere å oppdage, og det er ofte denne kundegruppen selskapene ønsker å identifisere og bevare. (Hadden et al 2007).

Hadden *et al.* (2007) forklarer at en kunde som tar en tilfeldig avgjørelse om å si opp kundeforholdet er ofte styrt av økonomiske og geografiske forhold som hindrer kunden i å fortsette kundeforholdet. Eksempler på slike forhold er henholdsvis at kunden ikke har råd til å fortsette kundeforholdet eller at kunden flytter til et sted der tjenesten ikke er tilgjengelig. (Hadden et al, 2007). Derimot oppstår ufrivillig kundefrafall som følge av at kunden nærmest blir «tvunget» til å forlate selskapet. Dette kan være et resultat av at kunden ikke betaler for tjenesten eller misbruker tjenesten, og dermed mister retten til å benytte seg av tjenesten (Hadden et al 2007, Olle & Cai, 2014). Denne kundegruppen er enklere å identifisere i motsetning til de som frivillig forlater selskapet (Hadden et al 2007).

3.2. Kostnader knyttet til kundefrafall

Det regnes som en bedre markedsføringsstrategi å bevare selskapets eksisterende kunder fremfor å tilegne seg nye kunder (Hadden et al ,2007; Kim et al, 2004). Det skyldes blant annet at det er en rekke kostnader forbundet med kundefrafall. Dette omfatter blant annet alternativkostnader som følge av manglende salgsinntekter og kostnader ved å tiltrekke seg nye kunder. For å tiltrekke seg nye kunder må selskaper ofte dekke kostnader til

promoteringer, rabatter, ressurser til å forstå kundenes behov og tid til å bygge et langvarige kundeforhold (Athanasopoulos, 2000). Med blant annet dette i betraktning har det vist seg at det å tiltrekke seg nye kunder kan være fem til seks ganger mer kostbart enn å beholde eksisterende kunder (Verbeke et al, 2012). I tillegg har studier vist at det å beholde eksisterende kunder kan gi økt profitt på lenger sikt. Reichheld & Sasser (1990) skriver at en reduksjon i kundefrafall på 5% kan skape opptil 85% mer i profitt.

3.3. Kundefrafall i telekommunikasjonsbransjen

I telekommunikasjonsbransjen ansees kundefrafall som en utfordring (Ahn et al. 2006; Al-Mashraie et al 2020). Årsaken til dette er delvis på grunn av utviklingen i teknologien som gjør det mulig for flere aktører å tilby mer varierte tjenester enn tidligere (Seo et al., 2008). I tillegg skyldes det økt konkurranse som følge av at antall bedrifter innen telekommunikasjonsbransjen har økt, og derfor er det lettere for kunder å bytte til en konkurrent dersom de er misfornøyde med leverandøren (Oskarsdottir et al., 2017).

Ifølge Al-Mashraie, et al. (2020) er det en sammenheng mellom selskapets «churn rate» og markedsandel (Al-Mashraie et al., 2020). Churn raten indikerer forholdet mellom kunder som forlater selskapet og antall kunder selskapet har i en periode (Ahn et al, 2020). I studiet ble det sett på churn raten og markedsandelen (basert på andel kunder) til fire amerikanske selskaper i telekommunikasjonsbransjen. Hver av selskapene hadde en gjennomsnittlig månedlig churn rate på 2,9%, 3.12%, 1,59% og 1,36% i 2014. I fjerde kvartal av 2014 var markedsandelen for disse selskapene på 16%, 15%, 34% og 34%. Det viser seg at markedsandelen er omvendt proporsjonal med selskapets «churn rate» (Al- Mashraie et al. 2020). Altså jo lavere churn rate, desto høyere markedsandel.

3.4. Maskinlæring

Maskinlæring handler om å trene opp algoritmer til å finne mønstre i store datasett for å ta beslutninger og prediksjoner basert på dette. Det er en undergruppe av kunstig intelligens (SAP, u.å.). Maskinlæring kan deles inn i *veiledet læring*, *ikke-veiledet læring* og *forsterket læring*. Veiledet læring innebærer å trene opp modeller basert på treningsdata med datainput

og ønsket resultat som er gitt på forhånd. Ikke-veiledet læring innebærer å identifisere strukturer og ekstrahere nødvendig informasjon fra et datasett uten forhåndsdefinerte svar. Forsterket læring legger vekt på å trene opp en modell som lærer og forbedrer seg gjennom tilbakemeldinger fra omgivelsene. (Raschka & Mirjalili, 2019).

I denne oppgaven anvendes det modeller inne veiledet og ikke-veiledet læring. Modellene innen veiledet læring som anvendes er *Decision Tree*, *Random Forest* og *Support Vector Machine*. I tillegg anvendes *Isolation Forest* som er en metode innen ikke-veiledet læring. Nedenfor presenteres teori knyttet til disse modellene.

3.4.1. Decision Tree

Decision Tree (Beslutningstre, DT) er en maskinlæringsmodell som kan sees på som en hierarkisk trestruktur som består av en rot-node, grener, interne-noder og en løv-node (sluttnode) (IBM, u.å.). Ved å starte fra rot-noden separeres observasjonene fra hverandre basert på den variabelen som maksimerer målfunksjonen, *Information Gain*. Denne splittingen skjer frem til alle observasjoner i hver sluttnode tilhører samme klasse. Det finnes ulike mål som brukes til å beregne hvor «rene» hver node er. Disse kalles *impurity measures* og noen eksempler på impurity-mål er *Gini impurity* og *Entropy*. (Raschka & Mirjalili, 2019).

Fordelen med Decision Tree er at den har en iboende tolkbarhet. Dette gjør det mulig å forstå hvordan modellen predikerer et utfall. Derimot er det økt sannsynlighet for overtilpassing med Decision Tree, på grunn av at observasjonene separeres helt til sluttnoden er «ren». Overtilpassing innebærer at den trente modellen ikke klarer å generalisere seg på urørt data. Ved å sett en grense på hvor mange splitt som skal gjennomføres, altså maksimum dybde på treer (*max_depth*) er det mulig å redusere overtilpassing. (Raschka & Mirjalili, 2019).

3.4.2. Random Forest

Random Forest baserer seg på prinsippene til *Decision Tree* (Beslutningstre), og er et ensemble av flere beslutningstrær. Ideen bak Random Forest er å ta gjennomsnittet av prediksjonen gjort av flere individuelle beslutningstrær. Den endelige prediksjonen i et

klassifiseringstilfelle med Random Forest vil være basert på en flertallsavstemning (*majority voting*). Random Forest algoritmen ansees som mer robust mot overtilpasning sammenlignet med en individuell beslutnings tre. En ulempe med Random Forest er at den ikke er like enkel å tolke som Decision Tree. (Raschka & Mirjalili, 2019).

Når det gjelder å finne den rette hyperparameteren med Random Forest, så trengs det i praksis å fokusere på antall trær. Jo flere beslutningstrær, desto bedre prediksjonsevne, men på bekostning av en økt beregningstid. (Raschka & Mirjalili, 2019).

3.4.3. Support Vector Machine

Algoritmen *Support Vector Machine* (Støttevektormaskin, SVM) separerer observasjoner som tilhører ulike klasser fra hverandre med en hyperplan som kalles for en beslutningsgrense (decision boundary). Denne algoritmen fokuserer på å maksimere avstanden (margin) mellom beslutningsgrensen og de observasjonene som befinner seg nærmest denne grensen. Observasjonen nærmest grensen kalles for støttevektorer. Hovedtanken bak SVM er å maksimere avstanden (margin) for å hindre at modellen overtilpasses. (Raschka & Mirjalili, 2019).

3.4.4. Isolation Forest

Isolation Forest er en ikke-veiledet metode innen maskinlæring som anvendes for å undersøke outliers og anomalier i et datasett (Gillespie, 2019). Outliers eller ekstremverdier er observasjoner som ikke samsvarer med de resterende observasjonene i datasettet. Dette kan blant annet skyldes at de ligger langt utenfor det forventede verdiområdet. (s. 358, Han & Kamber, 2006).

Isolation Forest algoritmen «isolerer» observasjoner ved å tilfeldig velge ut en variabel og deretter tilfeldig velge ut en verdi som ligger mellom den høyeste og laveste observasjonen for variablene hvor observasjonene splittes (scikit learn, u.å). Dette gjentas til alle observasjoner er splittet. Hovedtanken bak dette er at observasjoner som kan være outliers, vil ha en kortere vei fra rot-noden av treet til den endelige noden (leaf node), og outliers vil da

bli separert tidlig i prosessen med færre splitter. Ved å ta gjennomsnittet av veilengden for hver observasjon, er det mulig å finne utliggere i datasettet. (Gillespie, 2019).

3.4.5. Avveining mellom *bias* og *varians*

En utfordring innen maskinlæring er å ta en avveining mellom *bias* og *varians*. Målet er å unngå høy bias og høy varians. Dersom modellen har høy varians innebærer det at modellen har en lavere generaliseringsevne, og har overtilpasset seg treningsdata. Dette kan blant annet skyldes at modellen er for kompleks som følge av at den har for mange parametere. En modell med høy bias regnes som en undertilpasset modell. Dette innebærer at modellen ikke er kompleks nok til å finne viktig mønstre i treningsdata, som videre fører til at dens prediksjonsevne på testsettet ikke blir optimal. (Raschka & Mirjalili, 2019).

3.4.6. Evalueringskriterier

I maskinlæring er det vanlig å anvende ulike evalueringskriterier for å vurdere modellen. Evalueringskriterier anvendes for å evaluere og veilede modellens læringsalgoritme (Phung, Bouzerdoum & Nguyen, 2009). Derfor er det viktig å velge kriterier som er passende for det aktuelle problemet. I et klassifiseringsproblem med en ujevn klassefordeling er det viktig å sørge for at evalueringskriteriene tar hensyn til klassen med færrest observasjoner. Det mest brukte evalueringskriteriet *accuracy* som måler modellens treffsikkerhet, vil i tilfeller med klasseubalanse føre til at minoritetsklassen vil ha en mindre påvirkning på treffsikkerheten sammenlignet med majoritetsklassen (Phung, Bouzerdoum & Nguyen, 2009).

Følgende evalueringskriterier regnes som blant de mest relevante evalueringskriteriene for et datasett med ujevn klassefordeling: presisjon (*precision*), *recall*, *F-score*, *ROC AUC score*. Noen av disse kan forklares basert på en *forvirringsmatrise*. (Phung, Bouzerdoum & Nguyen, 2009). En forvirringsmatrise gir en oversikt over antall *Sann positiv (SP)*, *Sann negativ (SN)*, *Falsk positiv (FP)* og *Falsk negativ (FN)* prediksjoner fra en klassifiseringsmodell (Raschka & Mirjalili, 2019).

Sann positiv (SP) er antall virkelige positive tilfeller som har blitt predikert som positive av modellen. Sann negativ (SN) er antall virkelige negative tilfeller som har blitt predikert som

negative av modellen. Falsk positiv (FP) er antall virkelig negative tilfeller som har blitt feilpredikert som positive. Falsk negativ (FN) er antall virkelige positive tilfeller som har blitt feilpredikert som negative. Presisjonen til en klassifiseringsmodell er andelen av positive prediksjoner fra modellen som virkelig er positive. Recall er andelen av virkelig positive tilfeller som har blitt predikert riktig av modellen. F1-score betraktes som en kombinasjon av presisjon og recall. (Phung, Bouzerdoum & Nguyen, 2009).

$$Presisjon (PS) = \frac{SP}{SP+FP} \quad (1)$$

$$Recall (RC) = \frac{SP}{SP + FN} \quad (2)$$

$$F1 = \frac{2 \times RC \times PC}{RC + PC} \quad (3)$$

ROC (Receiver operating characteristic) og AUC anvendes for å evaluere den overordnede ytelsen til modellen. ROC visualiseres som en kurve, som viser forholdet mellom sann positiv raten (SPR) og falsk positiv raten (FPR). Diagonalen i en ROC-kurve betraktes som en «tilfeldig gjetting», og dersom en modell befinner seg under denne diagonalen ansees den som å ikke være bedre enn tilfeldig gjetting. En optimal modell vil finne seg nærmere øverste venstre hjørne av grafen. I dette område er sann positiv- raten (SPR) lik 1 og falsk positiv- rate (FPR) lik 0. (Raschka & Mirjalili, 2019). AUC er området under ROC-kurven. AUC-score på 0.5 tilsvarer tilfeldig gjetting, mens AUC-score lik 1 tilsvarer perfekt modell (Marzban, 2004). AUC- score er mindre partisk til klassestørrelse og er dermed egnet som evalueringskriterie i tilfeller med ujevn klassefordeling. (Phung, Bouzerdoum & Nguyen, 2009).

$$SPR = \frac{SP}{FN + SP} \quad (4)$$

$$FPR = \frac{FP}{FP + SN} \quad (5)$$

3.5. Forklarbarhet i maskinlæring

Rudin *et al.* (2022) skriver at forklarbar maskinlæring innebærer at «modellen følger en domenespesifikk sett med begrensninger som gjør det enklere for mennesker å forstå dataen eller modellens prediksjoner» (Rudin *et al.*, 2022). Målet med forklarbarhet i maskinlæring er å forsikre at beslutninger som drives av algoritmer kan forklares til sluttbrukeren og andre interessenter i ikke- tekniske termer (Barocas *et al.*, 2018 referert i Adadi & Berrada, 2018). Hovedpoenget er å skape tillitt til maskinlæringsmodeller og gjøre dem mer transparente. Det legges vekt på å anvende teknikker som gir mer forklarlige modeller samtidig som det høye ytelsesnivået til modellene opprettholdes (Adadi & Berrada, 2018).

Bodria *et al.* (u.å) forklarer at komplekse maskinlæringsmodeller anvendes i beslutningstaking for å oppnå høyere treffprosent i prediksjonene fremfor tolkbare modeller. I en studie undersøker Ashoori & Weisz (2019) hvilke faktorer som påvirker tillitten til kunstig intelligensbaserte beslutningsprosesser. De finner blant annet at modellenes tolkbarhet har en vesentlig betydning for hvorvidt mennesker har tillitt til kunstig intelligensbaserte beslutninger. Det viste seg at det var mindre tillitt til svart boks modeller fremfor forklarbare modeller.

Rudin (2018) forklarer at en modell betraktes som en svart- boks modell fordi den enten består av en funksjon som er for komplisert for mennesker å forstå eller at den består av en funksjon som er proprietær. Dyplæringsmodeller som består av rekursive algoritmer, er et eksempel på modeller som er kompliserte og forstå. En modell regnes som proprietær som følge av at produsentene er bekymret for at modellene kan rekonstrueres, og dermed velger kompliserte modeller av hensyn til hemmelighold (Rudin *et al.*, 2018). Et eksempel på en proprietær modell er COMPAS modellen som anvendes i det amerikanske rettssystemet for avgjørelser om prøveløslatelse og kausjon. Denne modellen har blant annet blitt kritisert for å være rasistisk og partisk i beslutningstaking. (Rudin, 2018).

Å anvende forklarbare modeller kan være til fordel da de bidrar til å forebygge feil i beslutningsprosesser som støtter seg på maskinlæringsmodeller på grunn av den dypere

innsikten modellene gir. Det kan også bidra til å unngå begrunnelser for beslutninger som «algoritmen fikk meg til å gjøre det» (FAT, u.å.).

Ved å undersøke 381 studier innen forklarbar kunstig intelligens kommer Adadi & Berrada (2018) frem til hvorfor det er behov for forklarbarhet i maskinlæringsmodeller. De kommer frem til at det er nødvendig med forklarbare modeller for å **1) begrunne resultatene** for å forsikre at maskinlæring baserte beslutninger ikke blir tatt på feil grunnlag. Dette er viktig for å forsikre at utfallet av modellene og beslutningene basert på disse kan forklares. I tillegg bevise at disse er rettferdige og etiske, da dette er viktig for å skape tillitt. **2) oppnå forsterket kontroll** ved bedre forståelse av systemet og dermed være i stand til å raskt identifisere og korrigere feil. Dette vil gi økt kontroll over systemet. **3) Kontinuerlig forbedring av modellene.** Ved å forstå hvorfor modellene predikerer et spesifikt utfall, vil det være enklere å komme på innovative måter på å forbedre den. **4) Oppdage ny kunnskap.** Forklaringer vil være et middel for å samle informasjon og tilegne seg ny kunnskap. Forklarbare modeller kan bidra til å forstå nye strategier. (Adadi & Berrada, 2018)

3.5.1. Hvorfor er det nødvendig med forklarbarhet i kundefrafallsprediksjon?

Når det gjelder kundefrafall er det ikke kun viktig for selskaper å identifisere kunder som kan forlate selskapet, men også forstå modellene som anvendes. For en maskinlæringsmodell som anvendes til prediksjon av kundefrafall er både prediksjon og forklarbarhet to viktige funksjoner (Shaw *et al.*, 2001, sitert i Qi *et al.*, 2009). Prediksjon vil hjelpe selskapene med å finne de kundene som mulig vil forlate selskapet, mens modellens forklarbarhet vil bidra til å oppdage mønstre i kundeatferden og gi dypere innsikt i faktorer som kan påvirke kundefrafall. Dessuten vil en dypere forståelse av kundeatferden gjennom forklarbare modeller, gjøre selskapene i stand til å sette inn relevante tiltak, og dermed oppnå bedre ressursallokering for arbeidet med kundebevaring. (Chen & Li 2006, sitert i Qi *et al.*, 2008).

3.5.2. Metoder innen forklarbar maskinlæring

Det eksisterer ulike metoder for forklarbarhet. Det skilles mellom modeller som er forklarbare i seg selv og metoder som anvendes for å forklare svart- boks modeller. Disse metodene kalles for *post-hoc* metoder, og kan videre deles inn i *globale-* og *lokale* metoder. Dessuten skilles det også mellom *modell agnostiske* og *modell spesifikke* metoder. (Bodria *et al.* 2021).

Transparente modeller

Arrieta *et al.* (2020) forklarer at enkelte maskinlæringsmodeller har iboende egenskaper som gjør at de er enklere å forstå dem uten ytterligere teknikker. Slike modeller består av algoritmer som ikke er for kompliserte til å forstås av mennesker. Det vil si at alle deler av modellen inkludert input, parametere, beregninger, output og selveste prosessen modellen gjennomgår fra input til output er forståelig for mennesker. Modeller med disse egenskapene regnes som transparente maskinlæringsmodeller. (Arrieta *et al.*, 2020). Å anvende transparente maskinlæringsmodeller kan ansees som den enkleste metoden for å oppnå forklarbare modeller på (Molnar, 2022). Decision Tree og lineære modeller er eksempler på transparente modeller (Du, Liu & Hu., 2019; Arrieta *et al.*, 2020).

Post- hoc metoder

Molnar (2022) påpeker at det anvendes post- hoc metoder for å forstå kompliserte modeller, men enkelte post-hoc metoder kan også anvendes på transparente modeller. Post- hoc metoder anvendes for å analysere modellene etter trening (Molnar, 2022). Murdoch *et al.* (2019) skriver at post- hoc metoder for lokale forklaringer fokuserer på å forstå modellen på prediksjonsnivå, mens globale forklaringer fokuserer på å forstå modellen på datasetts nivå (Murdoch, *et al.*, 2019).

Lokale forklaringer legger vekt på å forstå årsaken til et spesifikt prediksjonsresultat (Bodria *et al.*, u.å.). Ved å fokusere på individuelle prediksjoner fra modellen, vil det være enklere å forstå hva modellen predikerer for en bestemt input og hvorfor prediksjonen blir som det blir. Dette går ut på å finne hvilke variabler og sammenhenger som fører til det spesifikke prediksjonsresultatet. (Molnar, 2022).

Globale forklaringer legger vekt på å forstå den helhetlige logikken bak modellen (Bodria *et al.*, u.å.). Målet med å anvende post hoc metoder for globale forklaringer er å kunne forstå hvordan modellen tar beslutninger basert på dens variabler og hver av komponentene modellen har lært som for eksempel vektene, andre parametere og strukturen (Molnar, 2022).

Modellagnostiske metoder og modellspesifikke metoder

I følge Bodria *et al.* (2021) anvendes modellagnostiske forklaringsmetoder for å tolke alle typer svart- boks modeller. Molnar (2022) forklarer at modellagnostiske metoder er mer

fleksible med tanke på at de kan benyttes på nærmest alle maskinlæringsmodeller. Adadi & Berrada (2018) påpeker at modellagnostiske metoder ofte er post-hoc metoder, og kan deles inn i lokale- og globale forklarbare modeller. Derimot anvendes modellspesifikke forklaringsmetoder for å tolke noen få spesifikke modeller (Adadi & Berrada, 2018).

I denne oppgaven vil det i hovedsak anvendes SHAP som er to modellagnostiske metoder.

3.5.3. SHAP – Shapley Additive Explanation

SHAP (Shapley Additive Explanation) er en metode presentert av Scott M. Lundberg og Si-In Lee for å forstå prediksjoner fra avanserte maskinlæringsmodeller, ved å tildele hver variabel en verdi som definerer dens betydning for en bestemt prediksjon (Lundberg & Lee, 2017). Metoden tar utgangspunkt i *Shapely- verdier* som ble introdusert av Shapley i 1953 innen spillteori for å tildele utbetalinger til spillere basert på deres bidrag til den totale utbetalingen, hvor spillere oppnår en profitt som følge av samarbeidet (Molnar, 2022).

Hver variabel bidrar ulikt til et prediksjonsresultat, og hver variabels betydning kan defineres gjennom Shapley- verdier. Dette kan beregnes gjennom følgende formel (Perez & Bajorath, 2020):

$$\phi_i = \frac{1}{|N|!} \sum_{S \subseteq N \setminus \{i\}} |S|! (|N| - |S| - 1)! [f(S \cup \{i\}) - f(S)] \quad (6)$$

$f(S)$ tilsvarer prediksjonsresultatet til en ML modell som skal forklares, ved å anvende en mengde S av variabler, og N er den totale mengden av variabler. Det endelige bidraget (Shapley- verdi) til en variabel i er ϕ_i .

En av fordelene med SHAP er at den kan anvendes i både globale og lokale sammenhenger. I tillegg er det en fordel at den globale tolkningen samsvarer med de lokale forklaringene. Den har også en alternativ og relativ effektiv metode for tre baserte algoritmer som heter TreeSHAP. Ulempene med TreeSHAP er at den benytter en betinget forventning istedenfor marginal forventning, som medfører at enkelte variabler som egentlig ikke har noen betydning på modellen kan ende opp med å få en SHAP verdi høyere enn 0. Det er større sannsynlighet for at dett kan skje, dersom det er korrelasjon mellom variabler. (Molnar, 2022).

3.5.4. LIME

LIME (Local Interpretable Model agnostic Explanation) er en metode utviklet av Ribeiro et al. for å forklare kompliserte maskinlæringsmodeller ved å anvende en enklere modell som til en viss grad kan gjenskape prediksjonene til en komplisert modell (Pérez & Bajorath, 2020). Den enkle modellen kalles for *surrogatmodell* og anvendes for å forstå individuelle prediksjoner (Molnar, 2022).

Hovedtanken bak LIME er at den generer en ny variant av datasett som ble anvendt av svart-boks modellen som den trener en surrogatmodell på. En surrogatmodell kan være en enkel forklarbar modell som for eksempel Decision Tree. Den nye modellen vil da kun være en tilnærming til den avanserte modellen på prediksjonsnivå og ikke en god gjenspeiling på globalt nivå. (Molnar, 2022).

En av fordelene med LIME er at bruker et avstandskriterium for å evaluere hvor godt surrogat modellen klarer å gjenskape prediksjonene til svart-boks modellen. Dette vil gi en ide på hvor godt surrogatmodellen kan forklare svart-boks modellens prediksjoner. Samtidig er det en ulempe at det ikke finnes noen definisjoner på hvor stor avstand som blir for mye. I tillegg kan avstanden mellom to nære observasjoner få ulike forklaringer om prosessen gjentas igjen. (Molnar, 2022).

4. Litteraturgjennomgang

I denne delen presenteres tidligere studier på kundefrafallsprediksjon og studier som har fokusert på anvendelse av metoder innen forklarbar maskinlæring innen kundefrafallsprediksjon.

4.1. Prediksjon av kundefrafall

Flere studier har sett på prediksjon av kundefrafall. Det har blant annet blitt anvendt ulike metoder innenfor statistikk og maskinlæring. I dette avsnittet vil hovedfokusset være på å

presentere noen av funnene og maskinlæringsmetodene for kundefrafallsprediksjon fra litteraturen.

4.1.1. Sammenligning av klassifiseringsmetoder

Vafeiadis *et al.* (2015) sammenlignet fem av de mest anvendte klassifiseringsmetodene for kundefrafall i telekommunikasjonsbransjen. Klassifiseringsmetodene de vurderte var Artificial Neural Networks (ANN), Decision Trees (DT), Support Vector Machines (SVM), Naïve Bayes classifiers (NB) og Logistic Regression classifiers (LR). Hver modell ble sammenlignet med og uten boosting. Resultatene fra studiet viste at uten boosting hadde ANN med læringsalgoritmen bakover propagering høyest treffprosent, etterfulgt av DT og SVM. Med boosting fikk modellene en signifikant forbedret nøyaktighet. Den boostede versjonen av SVM (Poly med AdaBoost) hadde den høyeste treffprosenten på 96,85%, og var den modellen som viste seg å være den beste klassifiseringsmodellen for det aktuelle problemet. (Vafeiadis *et al.* 2015)

Tabell 1. Basert på Vafeiadis *et al.*, (2015). Sammenligning av ytelse for ANN-BPM, SVM, DT med og uten boosting.

Klassifiseringsmetode	Treffprosent (%)		F- score	
	Uten boosting	Med Boosting	Uten boosting	Med Boosting
ANN - BPN	94,06	95,05	77,48	80,97
SVM – RBF	93,18	96,05	73,16	84,22
SVM - Poly	93,04	96,85	73,11	84,57
DT	94,15	95,09	77,04	83,87
NB	86,94	-	53,31	-
LR	87,94	-	14,46	-

4.1.2. Hybridmodeller

Rajamohamed & Manokaran (2018) har studert på clustering og ulike maskinlæringsalgoritmer for å hovedsakelig predikere kundefrafall blant kredittkort kunder i bank. De gjennomførte først en grov K-means clustering for å fordele kundene i klynger og deretter ble ulike modeller anvendt for å gjennomføre klassifiseringen. Følgende modeller ble testet ut: Support Vector Machine, Random Forest, Decision Tree, K-nearest Neighbour og

Naive Bayes. Deres resultat viste at en kombinasjon av grov K- means clustering og Support Vector Machine ga høyest treffprosent og lavere misklassifiseringsfeil. Denne hybridmodellen ble i tillegg testet på et datasett innen telekommunikasjonsbransjen som ga en treffprosent på 95,75%. (Rajamohamed & Manokaran 2018).

I en studie av Ullah *et. al* (2019) fokuserer de på å anvende ulike klassifiseringsmetoder og k-means clustering for å kunne identifisere kunder som kommer til å forlate selskapet og samtidig gi en oversikt over hvilke faktorer som påvirker disse kundene. Etter å ha testet en rekke modeller finner de ut at Random Forest har en høyere treffprosent på 88.63%.

4.2.Forklarbare maskinlæring innen kundefrafallsprediksjon

I motsetning til andelen av eksisterende studier innen kundefrafallsprediksjon er det færre studier som har fokusert på å anvende metoder innen forklarbar maskinlæring i kundefrafallsprediksjon.

Tekouabou *et al.* (2022) har anvendt SHAP verdier for å kunne få en global forståelse av Random Forest modellen som ble anvendt for kundefrafallsprediksjon i banksektoren. I dette studiet ble SHAP anvendt for å få en forståelse av hvilke variabler som har hatt betydning for prediksjonen. Diaz, Galan & Carrasco (2022) har sett på churn innen B2B. De har blant annet anvendt metoder innen LIME, SHAP og ELI5 for å analysere prediksjoner gjennomført av XGBoost. De konkluderer med at både lokal og globale metoder gir økt forståelse av hvordan modellen har kommet frem til ulike beslutninger.

5. Metode

I denne seksjonen vil metoder anvendt for å undersøke data nærmere, forbehandle data og metoder anvendt i modelleringsprosessen bli presentert.

5.1.Data

I denne seksjonen presenteres datasettet som benyttes i dette studiet. Datasettet representerer kundedata fra en aktør innen telekommunikasjonsbransjen i USA og er tilgjengelig gjennom nettstedet *Kaggle* (Kaggle, 2020). Det består av 4250 rader og 20 kolonner, og har ingen

manglede verdier. Datasettet inneholder 19 uavhengige variabler og en avhengig variabel. Det finnes både numeriske- og kategoriske variabler i datasettet. I tabell 2-6 er de opprinnelige variablene i datasettet presentert.

Tabell 2. Uavhengige variabler.

Kundeinformasjon.

Nr.	Variabelnavn	Type	Forklaring
1.	<i>state</i>	Kategorisk (Nominal)	Bokstavkode med to bokstaver som representerer den staten kunden bor i.
2.	<i>area_code</i>	Kategorisk (Nominal)	Tresifret kode; eks. <i>area_code_AAA</i> , hvor AAA er et tresifret retningsnummer.
3.	<i>account_length</i>	Kontinuerlig	Antall måneder kunden har vært i et kundeforhold hos aktøren.

Tabell 3. Uavhengige variabler.

Abonnementstype

Nr.	Variabelnavn	Type	Forklaring
4.	<i>international_plan</i>	Kategorisk (Binær)	Indikerer om kunden har et abonnement med internasjonal tale (yes/no).
5.	<i>voice_mail_plan</i>	Kategorisk (Binær)	Indikerer om kunden har et abonnement med talepost (yes/no).

Tabell 4. Uavhengige variabler.

Brukshistorikk

Nr.	Variabelnavn	Type	Forklaring
6.	<i>number_vmail_messages</i>	Kontinuerlig	Antall talemeldinger.
7.	<i>total_day_minutes</i>	Kontinuerlig	Totalt antall minutter brukt på anrop på dagen.

8.	<i>total_day_calls</i>	Kontinuerlig	Totalt antall dagsanrop.
9.	<i>total_eve_minutes</i>	Kontinuerlig	Totalt antall minutter brukt på anrop på kvelden.
10.	<i>total_eve_calls</i>	Kontinuerlig	Totalt antall kveldsanrop.
11.	<i>total_night_minutes</i>	Kontinuerlig	Totalt antall minutter brukt på anrop på natta.
12.	<i>total_night_calls</i>	Kontinuerlig	Totalt antall nattanrop.
13.	<i>total_intl_minutes</i>	Kontinuerlig	Totalt antall minutter med internasjonale samtaler.
14.	<i>total_intl_calls</i>	Kontinuerlig	Totalt antall internasjonaleanrop.
15.	<i>number_customer_service_calls</i>	Kontinuerlig	Totalt antall anrop til kundeservice.

Tabell 5. Uavhengige variabler. Betalingsinformasjon

Nr.	Variabelnavn	Type	Forklaring
16.	<i>total_day_charge</i>	Kontinuerlig	Totalkostnader på dagsanrop.
17.	<i>total_eve_charge</i>	Kontinuerlig	Totalkostnader på kveldsanrop.
18.	<i>total_night_charge</i>	Kontinuerlig	Totalkostnader på nattanrop.
19.	<i>total_intl_charge</i>	Kontinuerlig	Totalkostnader på internasjonaleanrop.

Tabell 6. Avhengig variabel.

Klassifisering

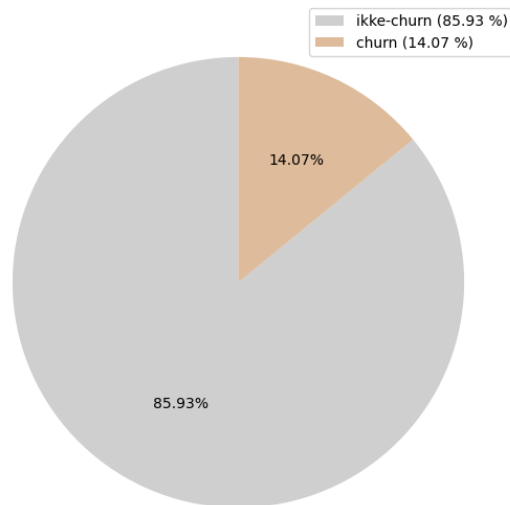
Nr.	Variabelnavn	Type	Forklaring
20.	<i>churn</i>	Kategorisk (Binær) Mål variabel	Indikerer kundefrafall (yes/no)

5.1.1. Ubalansert data

Datasettet består av en ujevn fordeling i den avhengige variabelen *churn*. Som vist i figur 1 finnes det 3652 observasjoner av totalt 4250 observasjoner som tilhører klassen «no».

Derimot er det kun 598 observasjoner av totalt 4250 observasjoner som tilhører klassen «yes».

Klassen «yes» er dermed overrepresentert med 85,93% i motsetning til klasse «no» som utgjør kun 14,07% av observasjonene.



Figur 1 Fordeling per klasse for den avhengige variabelen "churn".

Det er ikke uvanlig at antall kunder som forlater selskapet er betydelig mindre enn de som blir i selskapet, og dermed vil det føre til en ujevn klasse fordeling (Zhu, Baesens & Broucke, 2017). Dette vil skape problemer for de fleste klassifiseringsalgoritmer.

Et ubalansert datasett kan føre til utfordringer i modelltilpasningen. Et av problemene med et ubalansert datasett er at en ujevn klassefordeling medfører til at en maskinlæringsmodell trenes opp til at modellens beslutningsregel blir mer partisk mot den overrepresenterte klassen. Under modelltilpasningen gjennomføres det en optimering av en belønning- eller kostnadsfunksjon, som blir beregnet basert på summen av alle observasjonene i treningssettet. For å kunne minimere kostnaden eller maksimere belønningen, vil modellen dermed optimere prediksjonene basert på den overrepresenterte klassen i treningssettet. Et annet problem med ujevn klassefordeling er for eksempel å anvende mål som *accuracy* i modell evalueringen, da dette blir beregnet til fordel for den overrepresenterte klassen. (Raschka & Mirjalili, 2019).

5.2. Forbehandling av data

Det er kvaliteten på data og mengden av nyttig informasjon det inneholder som bestemmer hvor godt en maskinlæringsmodell kan lære. Dermed er det avgjørende å undersøke og forbehandle datasettet før selve modelleringsprosessen (Raschka & Mirjalili, 2019). I

forbehandlingsprosessen ble det fokusert på outliers og inkonsistens i datasettet samt håndtering av kategoriske variabler. Datasettet fra *Kaggle* hadde ingen manglete verdier, og dermed var det ikke behov for utfylling av disse.

5.2.1. Outliers

For å undersøke om det var outliers i de numeriske variablene ble det anvendt ulike metoder. Ved å visualisere data gjennom boksplokk og anvende metoden *Isolation Forest* ble det oppdaget at datasettet hadde mulige ekstremverdier. Ettersom outliers kan ha en negativ påvirkning på prediksjonsresultatet er det viktig å minimere effekten av outliers på modellen. I stedet for å fjerne disse verdiene fra datasettet ble de heller transformert til «akseptable verdier», slik det har blitt gjort i tidligere studier innen kundefrafallsprediksjon av Ali & Aritürk (2014) og Caigny *et al.* (2018). Denne metoden kalles for *Percentile Capping*. Ekstremverdier under 1. prosentil ble endret til samme verdi som observasjonen som befant seg i 1. prosentil og for de over 99. prosentil ble derfor endret til samme verdi som observasjonen som befant seg i 99. prosentil.

5.2.2. Korrelasjon

Ved å se på korrelasjonen for variablene i datasettet er det mulig å identifisere forholdet mellom de ulike variablene. Dersom det er høy korrelasjon mellom to variabler vil det føre til overflødig informasjon som kan føre til en overtilpasset modell (García, Luengo & Herrera, 2015). Korrelerte variabler kan gi misvisende SHAP verdier ved anvendelse av TreeSHAP. Ettersom datasettet inneholdt outliers ble det valgt å undersøke korrelasjonen mellom de numeriske variablene ved hjelp av *Spearman korrelasjon*. Spearman korrelasjon er en type korrelasjonskoeffisient som er mindre sensitiv mot outliers (Croux & Dehon, 2010; Akoglu, 2018). For de numeriske variablene som hadde en høy korrelasjonskoeffisient over 0.9, ble en av variablene fjernet. For å undersøke sammenhengen mellom de kategoriske variablene ble *Cramer's V* korrelasjon benyttet (Akoglu, 2018).

5.2.3. Håndtering av kategoriske variabler

I det opprinnelige datasettet finnes det totalt fem kategoriske variabler. Disse er de uavhengige variablene *state*, *area_code*, *international_plan*, *voice_mail_plan* og den avhengige variabelen *churn*.

Tabell 7. Oversikt over kategoriske variabler i datasettet.

Variabelnavn	Antall unike verdier	Utført behandling
<i>state</i>	51	fjernes
<i>area_code</i>	3	Genererer dummyvariabel
<i>international_plan</i>	2	Genererer dummyvariabel
<i>voice_mail_plan</i>	2	Genererer dummyvariabel
<i>churn</i>	2	Genererer dummyvariabel

De kategoriske variablene som har 2-3 unike verdier omgjøres til numeriske verdier ved å generere dummy variabler. Dette ble gjort ved hjelp av *get_dummies* funksjonen i Python. For å unngå problemer med multikollinearitet ble en av kolonnene som var omgjort til dummy variabel for hver kategorisk variabel fjernet. Det vil si at for variabelen *area_code* som i utgangspunktet blir transformert til tre ulike dummyvariabler med de tre unike *area_code* verdiene, ble en av kolonnene fjernet, slik at det var kun to kolonner som omfattet *area_code*. Variabelen *state* består av 51 unike verdier, og dersom denne variabelen omgjøres ved å generere dummy variabler, kan det føre til at datasettet får 51 nye variabler; ett binært variabel for hver unik verdi. Dette vil bidra til høy kardinalitet som videre kan føre til lavere generaliseringsevne (Carneiro *et al.* 2022). Derfor fjernes variabelen *state* fra datasettet.

5.2.4. Skalering av data

Det er vanlig å skalere datasettet før trening av modellene i maskinlæring. Ved å skalere verdiene i datasettet vil modellen kunne lære raskere. En av metodene for å skalere datasettet er gjennom standardisering. Dette innebærer å transformere variablene slik at forventningsverdien til hver variabel blir sentret rundt null og hver variabel vil få en standardavvik lik 1, slik en normalfordeling (Raschka & Mirjalili, 2019). I Python kan dette gjøres gjennom metoden *StandardScaler*.

5.2.5. Valg av modell og hyperparameter

Det ble valgt å anvende Decision Tree, Random Forest og Support Vector Machine. Disse modellene ble valgt basert på funn fra tidligere litteratur som viste at de hadde en relativ høy prediksjonsevne på kundefrafallsdata. De optimale hyperparameterne for hver modell ble funnet ved hjelp av *GridSearchCV* i Python. For å unngå overtilpassing og sørge for at klassefordelingen i det opprinnelige datasettet ble bevart under tuningen, ble det anvendt *Stratified 5-fold kryssvalidering*. I tabell 8 er det en oversikt over hyperparametere som ble testet ut.

Tabell 8. Oversikt over valg av hyperparametere for tuning.

Modell	Optimal hyperparameter
Decision Tree	Criterion: 'gini'. max_depth: min: 10, maks: 300, steg:20
Random Forest	N_estimator : Min: 10, maks: 500, steg: 50 Max_depth: Min:10, maks: 300, steg: 20 Max_features: 'auto', 'sqrt'
Support Vector Machine	Kernel : 'rbf' C: [0.0001, 0.001, 0.01, 0.1, 1.0, 10.0, 100.0] Gamma: [0.0001, 0.001, 0.01, 0.1, 1.0, 10.0, 100.0]

5.2.6. Splitting av datasett

Det er vanlig å dele datasettet i trenings- og test sett, slik at modellen trenes på treningssettet og deretter testes på testsettet, for å evaluere modellens prediksjonsevne. Dette gjøres for å unngå at modellen overtilpasser seg treningssettet, som videre vil svekke modellens generaliseringsevne. For å finne en akseptabel avveining mellom bias og varians, og for en optimal evaluering av modellen kan det anvendes ulike teknikker innen *kryss-validering* (Raschka & Mirjalili, 2019).

Ettersom datasett som anvendes i denne oppgaven er svært ubalansert, vil det være mest passende å anvende metoden *Stratified K-fold*. Dette innebærer at datasettet deles inn i K like store deler som gjenspeiler den eksisterende klassefordelingen i datasett. K-1 av delene anvendes til trening og den siste delen til validering. Dette gjentas K ganger. (Geiler, Affeldt & Nadif, 2022). Det velges å anvende K= 5 for dette datasettet.

5.3.SHAP

Etter å ha evaluert modellene ble det både anvendt metoder inne SHAP for å se på variabelens betydning for Random Forest modellen. Med utgangspunkt i forvirringsmatrisen ble det valgt å se nærmere på et tilfelle som ble feilklassifisert. Det ble gjort på en prediksjon som var falsk positiv. Falsk positiv vil si at et tilfelle som egentlig ikke er kundefrafall, har blitt feilklassifisert som kundefrafall. SHAP verdiene ble beregnet ved å anvende *TreeSHAP* i Python, selv om det ble påpekt i 3.3.3. at *TreeSHAP* kan gi misvisende resultater, blir det valgt å anvende denne metoden på grunn av dens effektivitet.

6. Resultat

I denne seksjonen presenteres resultatene fra modelleringsprosessen. Først blir resultater fra modelleringsprosessen presentert og deretter resultater etter fra av SHAP-analysen presentert.

6.1. Optimale hyperparameter

Etter hyperparametertuningen ble verdiene presentert i tabell 9 valgt ut som optimale basert på F1-score.

Tabell 9. Oversikt over de optimale parameterne.

Modell	Optimal hyperparameter
Decision Tree	Criterion: 'gini'. max_depth: 10
Random Forest	N_estimator : 30 Max_depth: 360 Max_features: 'auto'

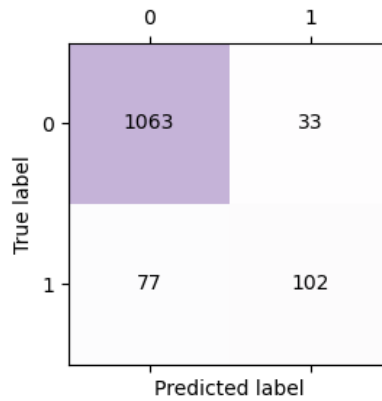
Support	Kernel : 'rbf'
Vector	C: 100
Machine	Gamma: 0.01

6.2. Forvirringsmatrise

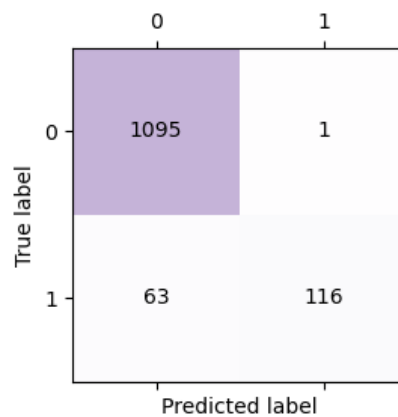
Etter å ha testet modellene på test settet ble modellene sammenlignet blant annet ved hjelp av en forvirringsmatrise. Matrisene for hver modell er presentert i figur 2, 3, 4, og gir en oversikt over modellenes klassifiseringsevne. Det kommer tydelig frem at enkelte observasjoner har blitt predikert feil. Når det gjelder antall tilfeller som har blitt feilpredikert som positiv (FP) skiller Random Forest (RF) seg ut, med kun et tilfelle med FP. For antall prediksjoner med falsk negativ (FN) er det ikke vesentlige forskjeller mellom modellene, bortsett fra at Decision Tree har færre tilfeller med FN og Support Vector Machine har flest tilfeller med FN som er lik 77. Når det gjelder korrekte prediksjoner, altså sann positiv (SP) og sann negativ (SN) viser det seg at SVM har et lavere antall sammenlignet med de andre.

	0	1
0	1067	29
1	57	122
	Predicted label	

Figur 2 Forvirringsmatrise for Decision Tree



Figur 3 Forvirringsmatrise for Support Vector Machine



Figur 4 Forvirringsmatrise for Random Forest

6.2.3. Evaluering

For å få et helhetlig bilde av klassifiseringsevnen til modellene ble ulike evalueringskriterier sammenlignet. I tabell 10 finnes det en oversikt over modellenes F1-score og AUC-score. AUC scoren for alle modellene ligger over 0.5, som tyder på at modellene har en relativt god klassifiseringsevne. Decision Tree og Random Forest scorer høyest basert på AUC score. Slik nevnt under 3.2.6. er AUC- score mindre partisk til klassestørrelser i et ubalansert datasett (Phung, Bouzerdoum & Nguyen (2009).

Tabell 10. Oversikt over de optimale parameterne.

	F1-score	AUC
DT	0.85	0.828
RF	0.80	0.823

SVM	0.87	0.769
------------	------	-------

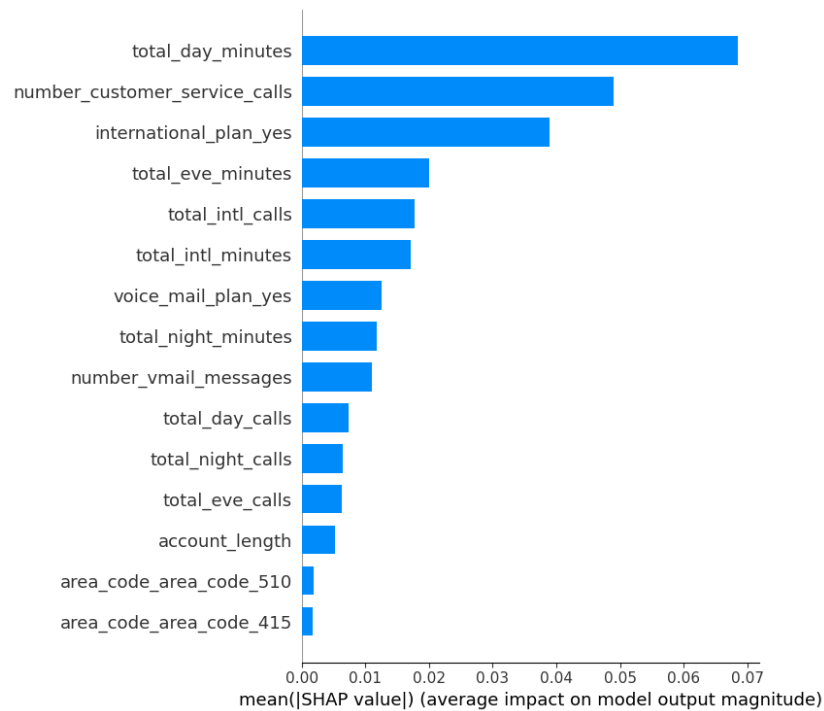
Slik det ble påpekt i 3.3.2. er Decision Tree en forklarbar modell. Fra resultatene er det mulig å se at en forklarbar modell scorer like høyt som mer avanserte maskinlæringsmodeller. I og med at hovedpoenget med denne oppgaven er å kunne finne ut om metoder innen forklarbar maskinlæring kan gi økt forståelse av en modell for kundefrafallsprediksjon, vil det legges mindre vekt på å finne den perfekte modellen. Derimot fokuseres det på å forstå modellene ved å anvende metoder innen forklarbar maskinlæring. Det ble derfor valgt å anvende disse metodene på Random Forest modellen.

6.3.SHAP for global forståelse av modellen

For å forstå hvilke variabler som har størst innvirkning på modellen ble det valgt å anvende SHAP. Dette ble gjort på Random Forest ved å anvende Tree SHAP i Python.

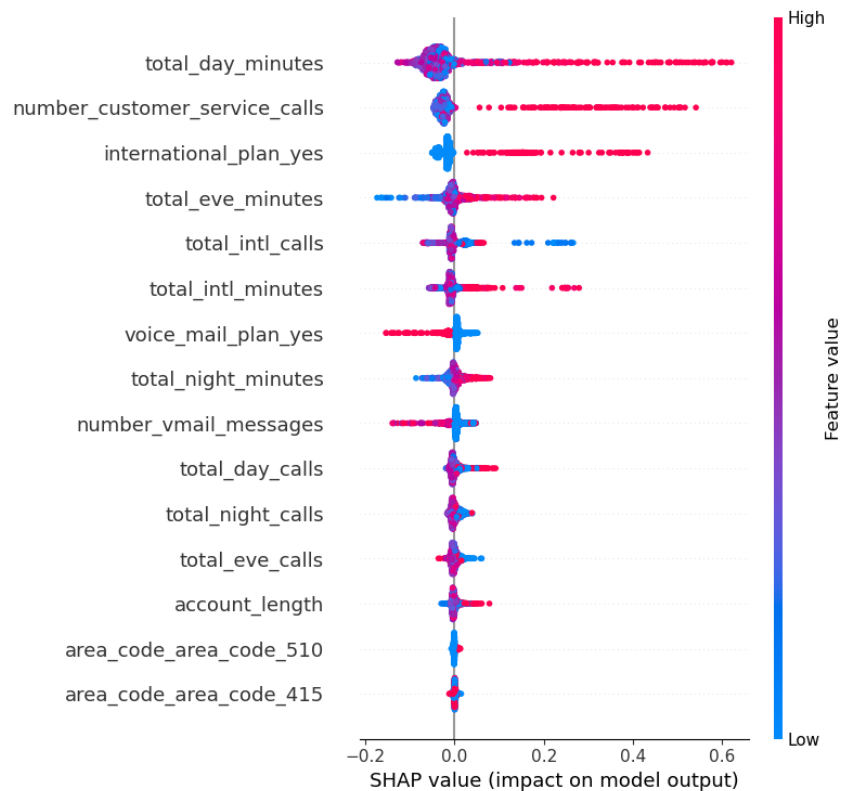
6.3.1. Variablenes betydning for modellens prediksjoner

I figur 5 er det en oversikt over et globalt barplott for variablene i datasettet som ble anvendt. Et globalt barplott gir en oversikt over variablenes betydning på globalt nivå, hvor betydningen av en variabel er gjennomsnittlig absoluttverdi av SHAP verdien for den variabelen over alle observasjoner i datasettet (Lundberg, 2018). Fra figur 5 er det mulig å se at variabelen *total_day_minutes*, *number_of_customer_service_calls* og *international_plan_yes* har størst betydning for modellen, mens *area_code_510* og *area_code_415* har mindre betydning. Utenom dette er det ikke mulig å trekke noen ytterligere slutninger fra modellen.



Figur 5 Globalt barplott som rangerer variablene etter gjennomsnittlig absoluttverdi av SHAP verdi for hver variabel over hele datasettet.

SHAP's sammendragplott i figur 6 viser en oversikt over variablenes betydning på modellen i sammenheng med hvilken effekt hver variabel har på modellen. Alle observasjoner for hver variabel i datasettet er fordelt som punkter i selve plottet. Hvert punkt er fordelt med utgangspunkt i sin egen SHAP- verdi som vises langs X-aksen. Y-aksen viser den virkelige verdien for hvert enkelt datapunkt i en variabel. Hvert datapunkt har en farge basert på sin virkelige verdi, der rød tilsvarer en høy verdi og blå tilsvarer en lav verdi. Områder i grafen der flere punkter befinner seg i klumper representerer klynger (Cooper, 2021).



Figur 6 Globalt sammendragplott av variablenes betydning.

Dersom den virkelige verdien til en observasjon i variabelen *total_day_minutes* er høy, vil variabelen få en positiv SHAP verdi. Det samme gjelder for *number_customer_service_call* og *international_plan_yes*. For disse variablene er det noen observasjoner som befinner seg i klynger. For *international_plan_yes* består denne klyngen av lave verdier som befinner seg i negativ retning, som vil si at de som ikke har et abonnement på *international_plan* bidrar negativt til kundefrafallsprediksjon. Dette innebærer at de har mindre sjanse for å forlate selskapet. Kunder som har en høy verdi for *total_day_minutes*, *number_customer_service_call*, *international_plan_yes*, *total_eve_minutes* har økt sjanse for frafall.

For *voice_mail_plan_yes* og *number_voice_mail_messages* er det omvendt, da høye verdier har en negativ SHAP verdi. Det vil si at kunder som har abonnement med talepost og kunder som har høyt antall talemeldinger har større sannsynlighet for å bli selskapet. Samtidig er det mulig se at kunder som har en kortere *account_length* vil mest sannsynlig bli i selskapet fremfor kunder med litt lengre *account length*. I motsetning til alle de andre variablene har

area_code_510 eller *area_code_415* nærmest ingen innvirkning på kundefrafallsprediksjoner for denne modellen.

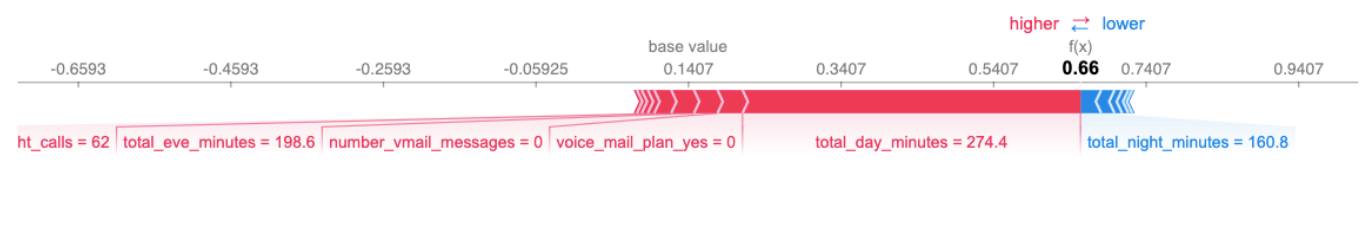
6.4. SHAP for lokal forståelse av modellen

Fra forvirringsmatrisen i 6.2. var det mulig å oppdage at Random Forest hadde klassifisert et tilfelle som feil positiv (FP) og totalt 63 tilfeller som feil negativ (FN). Det ble valgt å se nærmere på det ene tilfelle med FP, ved hjelp av SHAP.

6.4.1. Feil positiv (FP) – SHAP

Ved å sammenligne prediksjonene fra Random Forest modellen opp mot den virkelige klassen ble det funnet ut at observasjon nummer 187 egentlig ikke er et kundefrafallstilfelle, men ble feilaktig klassifisert som kundefrafall. Denne observasjonen ble undersøkt nærmere ved hjelp av SHAPs Forceplott (figur 7).

I en forceplott kan hver variabel forstås som en kraft som enten øker eller reduserer prediksjonen. Prediksjonene vil starte fra base- verdi som er gjennomsnittet av alle prediksjoner. For dette datasettet er gjennomsnittet for prediksjonene lik 0.1407 for klassen churn. For observasjon nummer 187 ble det predikert en verdi på 0.66. Dermed ble det antatt at det var 66% sannsynlighet for kundefrafall. I dette tilfellet er variablene *total_night_calls*, *total_eve_minutes*, *number_vmail_messages* og *voice_mail_plan_yes* med på å forskyve prediksjonen mot positiv retning, som vil si kundefrafall, altså churn = 1.



Figur 7 Forceplot av observasjon nummer 187 i datasettet.

6.4.2. Enkelttilfelle med kundefrafall på prediksjonsnivå

I figur 9 er det presentert en force plott med SHAP verdier for observasjon nummer 66 som er blitt riktig predikert som kundefrafall. Dette viser at `international_plan_yes` og `number_customer_service_call`, som også hadde betydning for modellen globalt, har påvirket til kundefrafall for denne observasjonen. Med en svært høy sannsynlighet på 99%.



Figur 8 Observasjon 66 som er riktig klassifisert som kundefrafall.

7. Diskusjon

Prediksjonsresultat

I denne oppgaven ble det undersøkt om hvorvidt metoder inne forklarbar maskinlæring kan bidra til å gi økt forståelse av kundefrafallsprediksjon i telekommunikasjonsbransjen. Ved å trene opp maskinlæringsmodeller som tidligere har visst seg å ha høy prediksjonsevne på kundefrafallsdata innen telekommunikasjonsbransjen ble det funnet ut i likhet med tidligere studier (*Vafeiadis et al., 2015*), at en forklarbar modell som Decision Tree kunne score like bra som avanserte modeller som SVM. Det at Decision Tree scoret omtrent like høyt som Random Forest kan skyldes at det ble satt en lav verdi for hyperparameteren `max_depth` for Decision Tree som kan ha virket mot overfitting.

Betydningen av variabler for kundefrafallsprediksjon

Det aktuelle datasettet som ble anvendt i denne oppgaven har relativt få variabler. Ved å gjennomføre en SHAP analyse på Random Forest modellen ble det funnet ut at av alle variablene var det variablene `total_day_minutes`, `number_of_customer_service_calls` og `international_plan_yes` hadde størst innvirkning på kundefrafallsmodellens prediksjon. At

antall telefoner til kundeservice har innvirkning på kundefrafallsmodellen, kan forstå ved at misfornøyde kunder kan kontakte kundeservice oftere. Det ble også oppdaget at *area_code_510* og *area_code_415* ikke var av noen betydning for modellens prediksjon en SHAP verdi nær null.

I tillegg ble det oppdaget at av de variablene som hadde høy score på den globale analysen av variablenes innvirkning viste seg også ha stor innvirkning på observasjon nummer 66 som ble riktig predikert som kundefrafall. For observasjon nummer 187 som ble feilpredikert som kundefrafall viste det seg at kun *total_eve_minutes* som hadde høy SHAP verdi i den globale observasjonen, var med på påvirke observasjon 187 som kundefrafall.

Datakvalitet

Datasettet som ble anvendt i dette studiet var hentet fra nettstedet *Kaggle* og hadde 19 uavhengige variabler og en avhengig variabel. Ettersom det ikke var noen grundig informasjon om datasettets opphav, er det ikke mulig å anta om dette her er reelle tilfeller. Dette kan ha hatt innvirkning på resultatet. Datasettet ble likevel anvendt på mangel på reelt datasett.

8. Konklusjon

Hovedpoenget med denne oppgaven var å undersøke om metoder innen forklarbar maskinlæring vil gi dypere i forståelse av kundefrafallsprediksjon i telekommunikasjonsbransjen. Det ble først fokusert på databehandling og modellering, før det ble anvendt metoder innen forklarbar maskinlæring.

Formålet med denne oppgaven var å undersøke følgende problemstilling:

I hvilken grad kan metoder innen forklarbar maskinlæring bidra til økt forståelse av kundefrafallsprediksjon i telekommunikasjonsbransjen?

Det ble satt opp to forskningsspørsmål for å undersøke problemstillingen nærmere.

- 1) *Hvordan kan globale modell-agnostiske metoder innen forklarbar maskinlæring bidra til en bredere forståelse av hvilke faktorer som påvirker en maskinlæringsmodell innen kundefrafallsprediksjon?*

I forhold til første forskningsspørsmål viste det seg at SHAP gir forståelse av hvordan ulike variabler påvirker modellen og prediksjonene på globalt nivå. Dette gir en ide om hvilke faktorer som kan ha en betydning når det gjelder kundefrafallsprediksjon og kan gi en viss innsikt til hvilke kunder som kan antas som potensielle churnere. I tillegg er det viktig å påpeke at dette kan være avhengig av hvilke variabler som befinner seg i datasettet. I tillegg ble det anvendt Tree-Shap som kan ha ført til mindre optimale resultater.

- 2) *Hvordan kan lokale modell-agnostiske metoder innen forklarbar maskinlæring bidra til en bredere forståelse av hvilke faktorer som påvirker individuelle prediksjoner?*

Når det gjelder de individuelle prediksjonene viser det seg at de nødvendigvis ikke er påvirket av de samme variablene i like grad som den overordnede modellen. Dette kan skyldes at de individuelle prediksjonene enten har høy eller lav verdi for en variabel og at dette kan ha gitt ulike resultater. Alt i alt så har analyser gjennom SHAP gitt et innblikk i hvilke faktorer som kan påvirker kundefrafall.

Videre forskning i dette område kan være å vurdere om kundefrafallsprediksjoner av mer avanserte modeller kan gi dypere innsikt ved hjelp av metoder innen forklarbar maskinlæring.

Referanser

- Adadi, A. & Berrada, M. (2018). Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI). *IEEE Access*, 6: 52138-52160.
doi.org/10.1109/ACCESS.2018.2870052.
- Ahn, J., Hwang, J., Kim, D., Choi, H. & Kang, S. (2020). A Survey on Churn Analysis in Various Business Domains. *IEEE Access*, 8: 220816-220839.
doi.org/10.1109/ACCESS.2020.3042657.
- Ahn, J., H. S. & Lee, Y. (2006). Customer churn analysis: Churn determinants and mediation effects of partial defection in the Korean mobile telecommunications service industry. *Telecommunications Policy*, 30 (10–11): 552-568.
doi.org/10.1016/j.telpol.2006.09.006.
- Akoglu, H. (2018). User's guide to correlation coefficients. *Turkish Journal of Emergency Medicine*, 18 (3): 91-93. doi.org/10.1016/j.tjem.2018.08.001.
- Ali, Ö. & Aritürk, U. (2014). Dynamic churn prediction framework with more effective use of rare event data: The case of private banking, *Expert Systems with Applications*, 41 (17): 7889-7903. doi.org/10.1016/j.eswa.2014.06.018.
- Al-Mashraie, M., Chung, S. H. & Jeon, H. W. (2020). Customer switching behavior analysis in the telecommunication industry via push-pull-mooring framework: A machine learning approach. *Computers & Industrial Engineering*, 144: 106476.
doi.org/10.1016/j.cie.2020.106476.
- Amin, A., Al-Obeidat, F., Shah, B., Adnan, A., Loo, J. & Anwar, S. (2019). Customer churn prediction in telecommunication industry using data certainty. *Journal of Business Research*, 94: 290-301. doi.org/10.1016/j.jbusres.2018.03.003.

- Arrieta, A. B., Díaz-Rodríguez, N., Ser, J. D., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R. & Herrera, F. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58: 82-115.
doi.org/10.1016/j.inffus.2019.12.012.
- Ashoori, M. & Weisz, J. D. (2019). In AI We Trust? Factors That Influence Trustworthiness of AI-infused Decision-Making Processes. *arXiv*. doi.org/10.48550/arXiv.1912.02675.
- Athanassopoulos, A. D. (2000). Customer Satisfaction Cues To Support Market Segmentation and Explain Switching Behavior. *Journal of Business Research*, 47 (3): 191-207.
doi.org/10.1016/S0148-2963(98)00060-5.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45: 5-32.
doi.org/10.1023/A:1010933404324.
- Bhattacharyya, J. & Dash, M. K. (2021). What Do We Know About Customer Churn Behaviour in the Telecommunication Industry? A Bibliometric Analysis of Research Trends, 1985–2019. *FIIB Business Review*, 11 (3): 280-302.
doi.org/10.1177/23197145211062687.
- Bodria, F., Giannotti, F., Guidotti, R., Naretto, F., Pedreschi, D. & Rinzivillo, S. (2021). Benchmarking and Survey of Explanation Methods for Black Box Models. *arXiv*.
<https://arxiv.org/abs/2102.13076>
- Caigny, A. D., & Coussement, K. & Bock, K. W. D. (2018). A new hybrid classification algorithm for customer churn prediction based on logistic regression and decision trees. *European Journal of Operational Research*, 269 (2): 760-772.
doi.org/10.1016/j.ejor.2018.02.009.

Carneiro, E.M., Forster, C.H.Q., Mialaret, L.F.S., Dias, L.A.V. & da Cunha, A.M. (2022). High-Cardinality Categorical Attributes and Credit Card Fraud Detection. *Mathematics*, 10 (20): 3808. doi.org/10.3390/math10203808.

Cooper, A. (2021). Explaining Machine Learning Models: A Non-Technical Guide to Interpreting SHAP Analyses. Tilgjengelig fra: <https://www.aidancooper.co.uk/a-non-technical-guide-to-interpreting-shap-analyses/> (lest 14.12.2022).

Croux, C. & Dehon, C. (2010). Influence functions of the Spearman and Kendall correlation measures. *Statistical Methods & Applications*, 19: 497-515. doi.org/10.1007/s10260-010-0142-z.

Deloitte. (2022). 2022 telecom industry outlook: Four trends pushing the industry forward Tilgjengelig fra: <https://www2.deloitte.com/us/en/pages/technology-media-and-telecommunications/articles/telecommunications-industry-outlook.html> (lest 14.12.2022).

Du, M., Liu, N. & Hu, X. (2019). Techniques for Interpretable Machine Learning. *arXiv*. doi.org/10.48550/arXiv.1808.00033.

FAT. (u. å.). *Bringing together a growing community of researchers and practitioners concerned with fairness, accountability, and transparency in machine learning*. Tilgjengelig fra: <https://www.fatml.org/> (lest 12.10.2022).

García, S., Luengo, J. & Herrera, F. (2015). *Data Preprocessing in Data Mining*. 1.utg. Cham: Springer. Tilgjengelig fra: <https://doi.org/10.1007/978-3-319-10247-4>.

Geiler, L., Affeldt, S. & Nadif, M. (2022). An effective strategy for churn prediction and customer profiling. *Data & Knowledge Engineering*, 142: 102100. doi.org/10.1016/j.datak.2022.102100.

- Gillespie, R. (2019). Detecting Fraud and Other Anomalies Using Isolation Forests. *Proceedings of SAS Global Forum*: 1-6. Tilgjengelig fra: <https://www.sas.com/content/dam/SAS/support/en/sas-global-forum-proceedings/2019/3306-2019.pdf>
- Gupta, S. & Lehmann, D. R. (2003). Customers as assets. *Journal of Interactive Marketing*, 17 (1): 9-24. doi.org/10.1002/dir.10045.
- Hadden, J., Tiwari, A., Roy, R. & Ruta, D. (2007). Computer assisted customer churn management: State-of-the-art and future trends. *Computers & Operations Research*, 34 (10): 2902-2917. doi.org/10.1016/j.cor.2005.11.007.
- Han, J & Kamber, M. (2006). *Data Mining: Concepts and Techniques*. 2.utg. Elsevier.
- Huang, B., Kechadi, M. T. & Buckley, B. (2012). Customer churn prediction in telecommunications. *Expert Systems with Applications*, 39 (1): 1414-1425. doi.org/10.1016/j.eswa.2011.08.024.
- IBM. (u. å.). *What is a Decision Tree? Learn the pros and cons of using decision trees for data mining and knowledge discovery tasks*. Tilgjengelig fra: <https://www.ibm.com/topics/decision-trees> (lest 13.12.2022).
- Jahanzeb, S. & Jabeen, S. (2007). Churn management in the telecom industry of Pakistan: A comparative study of Ufone and Telenor. *Journal of Database Marketing & Customer Strategy Management*, 14: 120–129. doi.org/10.1057/palgrave.dbm.3250043.
- Jahromi, A. T., Sepehri, M. M., Teimourpour, B. & Choobdar, S. (2010). Modeling customer churn in a non-contractual setting: the case of telecommunications service providers. *Journal of Strategic Marketing*, 18 (7): 587-598. doi.org/10.1080/0965254X.2010.529158.

- Kaggle. (2020). Customer Churn Prediction 2020: Predict whether a customer will change telco provider. Tilgjengelig fra: <https://www.kaggle.com/competitions/customer-churn-prediction-2020/data> (10.08.2022).
- Kim, M., Park, M. & Jeong, D. (2004). The effects of customer satisfaction and switching barrier on customer loyalty in Korean mobile telecommunication services. *Telecommunications Policy*, 28 (2): 145-159. doi.org/10.1016/j.telpol.2003.12.003.
- Lundberg, S. (2018). *SHAP: bar plot*. Tilgjengelig fra: https://shap.readthedocs.io/en/latest/example_notebooks/api_examples/plots/bar.html (lest 12.12.2022).
- Lundberg, & Lee. (2017). A Unified Approach to Interpreting Model Predictions. *Advances in neural information processing systems*, 30. [arXiv:1705.07874v2](https://arxiv.org/abs/1705.07874v2).
- Liu, F. T., Ting, K. M. & Zhou, Z. H. (2008). Isolation forest, *2008 Eighth IEEE International Conference on Data Mining*. Pisa, 15.-19. desember, 2008. IEEE, s. 413-422.
- Marzban, C. (2004). The ROC Curve and the Area under It as Performance Measure. *American Meteorological Society*, 19 (6): 1106-1114. doi.org/10.1175/825.1.
- Misra, R., Mahajan, V. & Mahajan, R. (2017). Review on factors affecting customer churn in telecom sector. *International Journal of Data Analysis Techniques and Strategies*, 2 (2): 122-144. doi.org/10.1504/IJDATS.2017.10006960.
- Molnar, C. (2022.). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable*. Tilgjengelig fra: <https://christophm.github.io/interpretable-ml-book/> .

- Murdoch, W. J., Singh, C., Kumbier, K., Abbasi-Asl, R. & Yi, B. (2019). Definitions, methods, and applications in interpretable machine learning. *Proceedings of the National Academy of Sciences*, 116 (44): 22071—22080. doi.org/10.1073/pnas.1900654116.
- Olle, G. D. O. & Cai, S. (2014). A Hybrid Churn Prediction Model in Mobile Telecommunication Industry. *International Journal of e-Education, e-Business, e-Management and e-Learning*, 4 (1): 55-62. doi.org/10.7763/IJEEEE.2014.V4.302.
- Óskarsdóttir, M., Bravo, C., Verbeke, W., Sarraute, C., Baesens, B. & Vanthienen, J. (2017). Social network analytics for churn prediction in telco: Model building, evaluation and network architecture. *Expert Systems with Applications*, 85: 204-220. doi.org/10.1016/j.eswa.2017.05.028.
- Pérez, R.R. & Bajorath J. (2020), Interpretation of machine learning models using shapley values: application to compound potency and multi-target activity predictions. *Journal of Computer-Aided Molecular Design*, 34 (10): 1013-1026. doi.org/10.1007/s10822-020-00314-0.
- Phung, S. L., Bouzerdoum, A. & Nguyen, G. H. (2009). Learning pattern classification tasks with imbalanced data sets. I: Yin, P. (red.), *Pattern Recognition*, s. 193-208. Rijeka: InTech. Tilgjengelig fra: doi.org/10.5772/7544.
- PwC. (u. å.). *Hva er kunstig intelligens?* Tilgjengelig fra: <https://www.pwc.no/no/teknologi-omstilling/digitalisering-pa-1-2-3/kunstig-intelligens.html> (15.11.2022).
- Qi, J., Zhang, L., Liu, Y., Li, L., Zhou, Y., Shen, Y., Liang, L. & Li, H. (2009). ADTreesLogit model for customer churn prediction. *Annals of Operations Research*, 168: 247–265 (2009). doi.org/10.1007/s10479-008-0400-8.

- Rajamohamed, R. & Manokaran, J. (2018). Improved credit card churn prediction based on rough clustering and supervised learning techniques. *Cluster Computing*, 21: 65–77. doi.org/10.1007/s10586-017-0933-1.
- Raschka, S. & Mirjalili, V. (2019). *Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow 2.3*. Utg. Mumbai: Packt Publishing.
- Reichheld F. F. & Sasser, W. E. (1990). Zero Defections: Quality Comes to Services. *Harvard business review*, 68 (5): 105-111.
- Richeldi, M. & Perrucci, A. (2002). Churn Analysis Case Study. *Deliverable D17*, 2. Tilgjengelig fra: https://www-ai.cs.tu-dortmund.de/PublicPublicationFiles/richeldi_perrucci_2002b.pdf.
- Rudin, C. (2018). Please Stop Explaining Black Box Models for High Stakes Decisions. *Stat*, 1050:26. Tilgjengelig fra: <https://www.datascienceassn.org/sites/default/files/Please%20Stop%20Explaining%20Black%20Box%20Models%20for%20High%20Stakes%20Decisions.pdf>
- Rudin, C., Chen, C., Chen, Z., Huang, H., Semenova, L. & Zhong, C. (2022). Interpretable machine learning: Fundamental principles and 10 grand challenges. *Statistics Surveys*, 16: 1-85. doi.org/10.1214/21-SS133.
- SAP. (u. å.). *What is machine learning?* Tilgjengelig fra: <https://www.sap.com/insights/what-is-machine-learning.html> (11.11.2022).
- scikit learn. (u. å.). *sklearn.ensemble.IsolationForest*. Tilgjengelig fra: <https://scikit-learn.org/stable/modules/generated/sklearn.ensemble.IsolationForest.html> (15.10.2022)
- Seo, D., Ranganathan, C. & Babad, Y. (2008). Two-level model of customer retention in the US mobile telecommunications service market. *Telecommunications Policy*, 32 (3–4): 182-196. doi.org/10.1016/j.telpol.2007.09.004.

- Stripling, E., Broucke, S. V., Antonio, K., Baesens, B. & Snoeck, M. (2018). Profit maximizing logistic model for customer churn prediction using genetic algorithms. *Swarm and Evolutionary Computation*, 40: 116-130. doi.org/10.1016/j.swevo.2017.10.010.
- Tékouabou, S.C.K., Gherghina, Ș.C., Touluni, H., Mata, P.N. & Martins, J.M. (2022). Towards Explainable Machine Learning for Bank Churn Prediction Using Data Balancing and Ensemble-Based Methods. *Mathematics*, 10(14): 2379. doi.org/10.3390/math10142379.
- Ullah, I., Raza, B., Malik, A.K., Imran, M., Islam, S. U. & Kim, S. W. (2019). A Churn Prediction Model Using Random Forest: Analysis of Machine Learning Techniques for Churn Prediction and Factor Identification in Telecom Sector. *IEEE Access*, 7: 60134-60149. doi.org/10.1109/ACCESS.2019.2914999.
- Vafeiadis, T., Diamantaras, K.I., Sarigiannidis, G. & Chatzisavvas, K. (2015). A comparison of machine learning techniques for customer churn prediction. *Simulation Modelling Practice and Theory*, 55: 1-9. doi.org/10.1016/j.simpat.2015.03.003.
- Verbeke, W., Dejaeger, K., Martens, D., Hur, J. & Baesens, B. (2012). New insights into churn prediction in the telecommunication sector: A profit driven data mining approach. *European Journal of Operational Research*, 218 (1): 211-229. doi.org/10.1016/j.ejor.2011.09.031.
- Zhu, B., Baesens, B. & Broucke, S. K. L. M. (2017). An empirical comparison of techniques for the class imbalance problem in churn prediction. *Information Sciences*, 408: 84-99. doi.org/10.1016/j.ins.2017.04.015.



Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitenskapelige universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway