Bacheloroppgave

NTNU
Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
NTNU Handelshøyskolen

Erlend Rake Ellingsen
Marthine Herger Lindgren

Kan en prediksjonsmodell i engelsk Premier League generere økonomisk fortjeneste?

AF3035 Bacheloroppgave i Business Analytics

Bacheloroppgave i Økonomi og Administrasjon
Veileder: Denis Becker
April 2024

NTNU
Kunnskap for en bedre verden

Erlend Rake Ellingsen
Marthine Herger Lindgren

Kan en prediksjonsmodell i engelsk Premier League generere økonomisk fortjeneste?

AF3035 Bacheloroppgave i Business Analytics

Bacheloroppgave i Økonomi og Administrasjon
Veileder: Denis Becker
April 2024

Norges teknisk-naturvitenskapelige universitet
Fakultet for økonomi
NTNU Handelshøyskolen

NTNU
Kunnskap for en bedre verden

Forord

Denne oppgaven representerer avslutningen på våre tre år ved NTNU Handelshøyskolen. Med hovedretning innenfor Business Analytics, gjenspeiler oppgaven vår interesse for analytiske verktøy og metoder. Gjennom arbeidet har vi fått et godt innblikk og forståelse i anvendelse av dataanalyseteknikker vi mener kommer til nytte i fremtiden.

I startfasen av vår bacheloroppgave var det mye diskusjon og usikkerhet knyttet til valg av tema og problemstilling. Gjennom god dialog med Denis Backer og Morten Kringstad, oppsto det et spennende tema for valg av oppgaven vår - betting. Dette temaet ga oss muligheten til å kombinere våre personlige interesser for idrett med forskningsmetoder. Kombinasjonen av dette inspirerte oss til å utforske fotballens komplekse bettingverden.

Avslutningsvis vil vi takke Dennis Backer for veiledning og hjelp, både før og under hele skriveprosessen. I tillegg vil vi takke Morten Kringstad for hjelp med utforming av oppgavens tema og problemstilling, samt god oppføling av oppgaven.

Innholdet i denne oppgaven står for forfatternes regning.

Erlend Rake Ellingsen

Marthine Herger Lindgren

Sammendrag

Formålet med denne oppgaven er å lage et verktøy som kan bidra til kontrollert betting. Basert på dette skal oppgaven konkret ta for seg hvordan en logistisk regresjonsmodell kan benyttes til å ta informerte bettingvalg, i tillegg til å vurdere modellens evne til å skape økonomisk fortjeneste.

Oppgaven tar utgangspunkt i kampresultater fra Premier League sesongen 2018/2019 for å predikere utfall og odds. Ut ifra den tilgjengelige dataen har vi valgt variabler som kan være relevant for modellen. Disse variablene er *form generelt*, *form på hjemme eller bortebane*, *resultater sist lagene møttes*, *resultat i ligaen forrige sesong* og *hjemmebanefordel*. Til å starte med blir modellen trent på 280 kamper, før den blir testet på sesongens resterende 72 kamper.

Resultatene viser en presisjon på 62,5% i predikering av kampresultater. Alt i alt konkluderer vi med at modellen er et godt verktøy for betting, da modellen bidrar til en overordnet gevinst på 12 501 kroner over 72 kamper på testsettet. Likevel har modellen noen svakheter, slik som å predikere uavgjort-utfall. Grunnen til dette kan være modellens begrensede antall variabler, eller idrettens kompleksitet. I utbredelse av modellen foreslår vi derfor å inkludere flere variabler for å kunne oppnå en enda bedre modell.

Abstract

The purpose of this study is to develop a tool that can help betting to be controlled. Therefore, this thesis will explore how a logistic regression model can be used to make informed decisions and evaluating whether the model will generate profit.

The basis of this thesis is match results from the English Premier League 2018/2019. From this dataset, variables are defined to predict full-time match results and odds. These variables include form in general, form based on home or away court, previous match between the respective teams, league results from last season and home court advantage. The model uses 80% of the data as training, and 20% to test the model, which constitutes 280 and 72 matches.

The results from the analysis showed that the model had an accuracy of 62,5%. Therefore, the conclusion is that the model can be used as a tool to perform controlled betting. Overall, the test set generated a profit of 12 501 kroner, which makes the model successful. In addition, the model doesn't predict any draw results, which express its weakness. Due to the limited number of variables or the complexity of the sport, the model is limited. To make the model even better, we suggest including more variables to capture the sport's complexity.

Innholdsfortegnelse

Figur- og tabelloversikt

1. Innledning

1.1 Bakgrunn for oppgaven

I dagens samfunn utvikler teknologien seg raskt, og i takt med dette øker tilgjengeligheten av betting- og gamblingmuligheter. I 2023 rapporterte Helsedirektoratet at omtrent 60 prosent av befolkningen i alderen 16-74 år drev med pengespill, en situasjon som kan føre til personlig økonomisk tap (Helsedirektoratet, 2022). Likevel kan gambling drives behersket dersom det praktiseres ansvarlig. Et sunt forhold til gambling innebærer at spillerne er velinformert og tar veloverveide beslutninger. Å informere spillere kan derfor bidra til bedre kontroll og mer ansvarlig spill.

Bettingmarkedet inkluderer mange ulike spilltyper, hvorav betting på fotballkamper er det mest utbredte (Complete Sports, 2024). Fotball er kjent som verdens mest populære idrett, noe som medfører en omfattende mediedekning (Lindner, 2024). Dette sikrer tilgjengelig data og statistikk som er avgjørende for utviklingen av analytiske modeller. Idrettens tydelig definerte sesonger og forutsigbare kampoppsett legger til rette for en systematisk tilnærming til datainnsamling og analyse. Engelske Premier League, som den mest populære ligaen innenfor fotball, er spesielt gunstig for slike analyser (Zamain, 2023). På grunn av idrettens oppsett og popularitet har vi valgt å predikere odds for Premier League kamper i denne oppgaven.

1.2 Problemstillingen

Konkret i denne oppgaven skal vi undersøke hvordan en logistisk regresjonsmodell kan benyttes til å ta informerte beslutninger om bettingvalg. Modellen anvender maskinlæring og statistisk analyse av historisk data fra Premier League kamper. Videre skal vi evaluere modellens evne til å generere økonomisk fortjeneste, ved å simulere innsatser basert på prediksjoner sammenlignet med faktiske kampresultater og ved bruk av reelle odds. På denne måten kan spillere drive betting med større grad av kontroll.

2. Teori

2.1 Generelt om betting

Betting går ut på å tippe på forventet utfall av en begivenhet. Ulike begivenheter kan være sjakk, virtuelle spill, e-sport, golf, eller fotball. I denne oppgaven vil betting dreie seg om betting på fotballkamper med utgangspunkt i Premier League sesongen 2018/2019.

Forventet utfall av en kamp uttrykkes vanligvis med en odds. Oddsen representerer sannsynligheten for tre ulike utfall: seier, tap eller uavgjort. I denne oppgaven skal vi forholde oss til odds på desimaltall. Med en lav odds vil dette reflektere en høy sannsynlighet for utfallet, og med en høy odds reflektere en lav sannsynlighet. Forholdet mellom odds og sannsynlighet vil være en likning hvor x representerer oddsen for hendelse i. i tar for seg utfallene hjemmeseier, borteseier eller uavgjort, hvor p representerer sannsynligheten for tilhørende utfall. Eksempelvis vil en lav odds på 1,56 gi en høy sannsynlighet på 64%, mens en høy odds på 5,56 vil gi en lav sannsynlighet på 18%.

$$x_i = \frac{1}{p}$$

Formel 1 Uttrykk for odds

2.2 Sannsynlighet for seier, tap eller uavgjort

En fotballkamp har tre ulike utfall som er hjemmeseier, uavgjort eller borteseier. Utfallet blir påvirket av en rekke faktorer, eksempelvis hjemmefordel eller form. Videre skal vi ta for oss sammenhengen mellom utfallene og predikering av dem i litteraturen.

Det er knyttet usikkerhet til prediksjon av kampresultater. Peel og Thomas (1992) viser til størst usikkerhet for et kampresultat når sannsynligheten for utfallene er like. Det vil si at sannsynligheten for de tre ulike utfallene er på 33,33%. I et slikt tilfellet vil sannsynligheten for alle utfall være like stor, og derfor størst usikkerhet.

Det er to forhåndsdefinerte faktorer som kan bidra til å forutsi vinneren i forkant av en kamp. For det første har forskning av Peel og Thomas (1992) vist en høy korrelasjon mellom tabellplassering og utfallet. Altså er det konstatert at lag høyt på tabellen har en tendens til å vinne over lag lengre ned på tabellen. For det andre er antall mål scoret i en kamp signifikant

korrelert med kampresultatet. Basert på 3914 kamper spilt over en 11 års periode, ble det konkludert med at laget på bortebane slipper inn flere mål og scorer færre mål enn lag på hjemmebane (Hoås, 2012). Det vil si at ved borteseier blir kampen i snitt vunnet med færre scorede mål enn ved hjemmeseier.

I artikkelen til Peel og Thomas (1992) fremhever de fastsatte odds som en presis prediksjon av kampresultater. Gjennom deres modell konkluderer de med at oddsen til bettingsselskap er en god modell i forhold til andre modeller i litteraturen. Dette er fordi bettingselskap har komplekse modeller som ofte inkluderer flere variabler og har oppdatert informasjon (Peel & Thomas, 1992).

I litteraturen til Hucaljuk og Rakipović (2011) er det brukt maskinlæring i forsøk på å predikere kampresultater. De kom frem til en modell med opp mot 60% treffsikkerhet på resultatene i Champions League. I deres konklusjon understreker de at fotballen er en kompleks idrett med mange variabler, slik som individuell form hos spillere. I tillegg til dette fremhever Hucaljuk og Rakipović (2011) vanskeligheten ved å kvantifisere kvalitative variabler. Dette viser utfordringer i utarbeidelsen av en modell for å predikere utfall (Hucaljuk & Rakipović, 2011).

2.3 Hjemmefordel

Å spille en kamp på hjemmebane er stor fordel for hjemmelaget. I litteraturen er Forrest og Simmons (2002) et eksempel på noen som har belyst dette temaet. I deres artikkel var sannsynligheten for borteseier høyere enn hjemmeseier i kun 72 av 872 tilfeller. Dette viser fordelen hjemmelaget har av å spille på hjemmebane (Forrest & Simmons, 2002).

Hjemmefordelen kan forklares av ulike faktorer. Reisevei til bortelaget og publikumsstøtte er eksempler på påvirkningsfaktorer til kampresultatet. Ifølge Peel og Thomas (1992) kan bortelagets reisevei påvirke spillernes prestasjon (Peel & Thomas, 1992). I Hoås sin masteroppgave fokuseres det på hvordan fysisk og psykisk påkjennelse fra reisevei påvirker prestasjonen. Fordelen av å spille på hjemmebane kan sees i sammenheng med kjennskap til underlag og banestørrelse. Stress trekkes også frem som en ytre faktor i forberedelsestiden, slik som å sove på hotell i stedet for i eget hjem før kamp. Å spille kamp på bortebane er altså

en stor påkjenning mentalt og fysisk, ettersom utøverne utsettes for nye forhold og inntrykk (Hoås, 2012).

Hjemmebanefordelen utnyttes på mange forskjellige områder, både av klubber og fans. En viktig faktor som blir planlagt til minste detalj er bortegarderoben. Det finnes flere ulike eksempler på hvordan forskjellige hjemmelag har designet bortegarderoben for å få en psykologisk fordel. I den engelske avisen The Sun trekkes det frem eksempler slik som å skru opp varmen i bortegarderoben, gjøre døra ut fra bortegarderoben større enn sin egen, polere gulvet for at det skal være glatt å gå med fotballsko og sette inn smalere speil enn vanlig (Morgan, 2018). Alt dette er små detaljer som er med på å psyke ut bortelaget før kamp.

Fansen driver psykologisk spill både før og etter kamp. Det har vært flere episoder der fans har angrepet bortelagets spillerbuss på vei til stadion (Jackson, 2023). Dette kan anses som dårlig oppførsel, men er et forsøk på å psyke ut motstanderen. Også underveis har hjemmelagets supportere en stor fordel og påvirkning. I de aller fleste tilfeller er hjemmelagets supportere i flertall i forhold til bortefansen. Dette gjør at hjemmefansen kan ha et høyere lydnivå, utøve pipekonsert mot motstanderen eller generelt skape en god atmosfære for hjemmelaget.

Over tid har hjemmefordelen sunket noe. Dobson og Goddard (2011) viser til en undersøkelse av fire divisjoner i engelsk fotball hvor hjemmefordelen har stagnert. Samlet sett har hjemmefordelen gått fra 64,4% i sesongene 1970-1974 til 57,7% i 2005-2009. Tema som diskuteres er utviklingen av teknologien og hvordan dette har gjort reisevei enklere og til en mindre belastning over tid. Det kan derfor diskuteres hvorvidt hjemmefordelen vil være like sterk i fremtiden. (Dobson & Goddard, 2011)

2.4 BET365

BET365 er et globalt bettingsselskap basert i England som ble etablert i år 2000. Siden oppstarten har selskapet opparbeidet seg 90 millioner kunder over hele verden (BET365, 2024). Gjennom sitt «one-wallet-system» får kundene tilgang til en rekke bettingformer, inkludert Sport, Casino, live-Casino, Games, Poker og Bingo.

BET365 henter fotballstatistikk gjennom en tredjepartsleverandør. Opta leverer statistikk som benytter en kombinasjon av AI-teknologi og menneskelig ekspertise. Statistikken de tilbyr er eksempelvis "Shot on Target", "Shot off Target", "Player Tackles" eller "Corners". Denne informasjonen bruker BET365 til å generere odds, som representeres for kundene i et live-format. Dette sikrer at kunden alltid er oppdatert på nyeste odds, i tillegg til å kunne hente ut relevant informasjon fra BET365 sin nettside. Oddsene fra BET365 reflekterer deres vurdering av sannsynligheten for ulike utfall, noe som vil si at usannsynlige utfall ikke er umulige (BET365, 2024).

2.5 Bettingselskap som risikoavers aktør

Innenfor økonomi- og beslutningsteori er risikoaversjon et sentralt begrep. Med en risikoavers aktør mener vi en aktør som unngår risiko (Stefánsson & Bradley, 2019). Basert på en persons individuelle tilnærming til nytte og sannsynlighet vil holdninger til risiko variere. I et tilfelle hvor en aktør står ovenfor valget om å investere 50 kroner for å kunne vinne 100 kroner eller risikere å tape pengene, vil en risikoavers aktør ikke gamble dersom man ikke kan forvente å vinne 50 kroner. Risikoen avhenger av forventet nytte av investeringsprisen og forventningsverdien på utfallet. For bettingselskap vil dette si at oddsen må kombinere sannsynligheten for et utfall og nytten (profitt) for å kunne drive lønnsomt (Stefánsson & Bradley, 2019).

De fastsatte oddsene til bettingselskap inkluderer fenomenet «Overround». Overround vil si at summen av oddsene overstiger 100%. Grunnen til dette er for å kompensere for påtatt risiko, dekke driftskostnader og sikre fortjeneste for bettingselskapene. Dette gjør at fastsatte odds ikke representerer forventet resultat fullstendig, da det er justert for profitt og risiko (Sestovic, 2017).

Basert på funksjonen av en forventningsverdi kan en «fair odds» defineres slik som i Formel 1. Ettersom at bettingselskap tar forbehold i sin odds, vil derfor den justerte oddsen være illustrert i Formel 2, hvor 1 + Ei er feilleddet eller justeringen som gjøres basert på risiko og profitt.

$$d_i^{\text{fair}} = 1/p_i.$$

Formel 2 Forventningsverdi for en fair odds.

$$d_i = \frac{d_i^{\text{fair}}}{1+\xi_i} = \frac{\frac{1}{p_i}}{1+\xi_i}.$$

Formel 3 Forventningsverdi for odds innlagt risiko og profitt

Sestovic (2017) konkluderer med at bettingselskap sine odds reflekterer motivasjon til å maksimere profitt og minimere risiko, fremfor å predikere mest nøyaktig resultat. Samtidig påpeker Sestovic en historisk nedgang i hvordan bettingselskapene justerer odds, noe som viser at selskapene har påtatt seg mer risiko de siste ti årene. I tillegg til dette utnytter de kognitive svakheter ved mennesket for å øke inntektene sine. Menneske er påvirket av mentale skjevheter som bettingselskap bruker for å tjene penger (Sestovic, 2017).

2.6 Kunder og investeringer

Med online plattformer og apper gjør dette kundene få tastetrykk unna å kunne bette på kamper. Tilgjengeligheten åpner opp for muligheter og et stort utvalg av scenarioer å bette på. Flere bettingselskap tilbyr for eksempel å bette på antall mål som scores i løpet av en kamp, hvilke spillere som kommer til å score, antall cornere eller hvilket lag som vinner kampen. Den økte tilgjengeligheten har imidlertid også ført til lavere terskel for å bette blant kundene. Dette setter krav til selvkontroll og personlig økonomi for å ha et kontrollert forhold til betting. Til tross for risikoen forbundet med spillavhengighet, forblir bettingmarkedet stort. Med dette som utgangspunkt vil vi følgelig undersøke motivasjonen til spillere og grunnene til at de velger å påta seg risiko (Jeannotsson & Ingvarsson, 2023).

I masteroppgaven til Henriksen og Nilsen (2020) blir nytteverdi eller nytteegenskaper brukt for å forklare motivasjon til gamblere. De tar for seg gjennomgående motiver for gambling, slik som å komme seg bort fra hverdagen, sosiale årsaker, utfordring, spenning, selvtillit, penger og drømmen om å vinne storpremien. Disse motivasjonene skaper en nytte gjennom opplevd belønning, hvor belønning eksempelvis kan være mestring eller penger (Henriksen & Nilsen, 2020).

Kultur og sosialt nettverk påvirker beslutninger. Av Jeannotsson og Ingvarsson (2023) fremkommer det at en kunde er mer tilbøyelig til å bette på et lag dersom venner eller familie gjør det samme (Jeannotsson & Ingvarsson, 2023). Mennesker blir påvirket av hverandre, og formes av normer og forventninger i samfunnet. Personer blir motivert av sosiale rammer og relasjoner, for eksempel kan en gambler bli motivert til å bette for å føle på en gruppetilhørighet. Fotball kan betraktes som et sosialt spill mellom ulike sosiale grupper, noe som forsterker tilhørigheten og konkurransen mellom ulike fan-grupper. Altså kan gamblere ta på seg risiko for å ta del av et sosialt nettverk eller føle på en gruppetilhørighet (Na, et al., 2019).

Demografiske faktorer påvirker også hvem som better. En studie gjort av Hing et al. (2016) viser at fellestrekk ved gamblere som påtar seg stor risiko er ung mann, singel, utdannet, full-tids jobb eller full-tids student. Det kan imidlertid tenkes at disse faktorene har sammenheng med individuelle faktorer (Hing, et al., 2016). Henriksen og Nilsen (2020) nevner ulike personlighetstrekk som beskriver gamblere. Dette er for eksempel mennesker som er optimistiske, risikotilbøyelige og overtroisk.

Michael et al. (2009) diskuterer «Favorite-Longshot Bias» som et eksempel på en skjevhet i spilleres vurdering av sannsynligheter. Dette er et fenomen der en spiller overestimerer sannsynligheten for et usannsynlig utfall. Det vil si at spilleren velger å investere i et tilfelle hvor sannsynligheten for utfallet er meget lav. Artikkelen trekker frem fellestrekk ved disse kundene, som for eksempel uinformerte og tilfeldige valg. Altså er longshoot bias den kognitive skjevhet som gjør at noen kunder velger å bette på tilnærmede umulige utfall, fordi kundene er uvitende eller sitter på begrenset med informasjon (Michael, et al., 2009).

I litteraturen blir «Fan-identity Based Bias» beskrevet som en skjevhet basert på lagtillit. Spillere er mer tilbøyelig til å bette på sitt favoritt-lag, fremfor andre lag, fordi spilleren sitter på subjektiv informasjon om laget. Gamblere samler inn lag-spesifikk informasjon for å underbygge troen på at man kan forutsi utfallet av en kamp. I kombinasjon med dette og gamblerens personlige mening blir derfor prediksjonen farget av deres lagtillitt. Kunkel (2018) trekker frem at sterkt involverte fans i større grad vil være påvirket av bias, sammenliknet med mindre involverte spillere (Na, et al., 2019).

3. Metode

3.1 Valg av data

Utgangspunktet i denne oppgaven er et datasett fra Premier League sesongen 2018/2019. Datasettet er den nyligst fullførte sesongen uten betydelig påvirkning fra koronapandemien. Påfølgende sesonger ble forstyrret av pandemirelaterte faktorer som kan ha påvirket resultatet, slik som tribunerestriksjoner. I tillegg har strenge reiseregler og økt fravær av spillere grunnet sykdom, ført til at senere sesongene avviker fra normale forhold. Derfor anser vi sesongen 2018/2019 som det mest representative utgangspunktet for en pålitelig modell.

3.2 Valg av variabler

Modellen bygger på sekundærdata fra Kaggle.com. Oppgaven vil primært bygge på resultatene fra sesongen 2018/2019, men vi vil også benytte noe data fra sesongen 2017/2018 for å supplere variablene og kampene som er spilt tidlig på sesongen. Datasettet består av totalt 380 kamper fordelt på 20 lag med 38 serierunder. Dette innebærer at hvert lag spiller mot hverandre to ganger, en kamp på hjemmebane og en på bortebane.

Basert på datasettet er det utarbeidet fem variabler som skal brukes for å predikere sannsynlighetene for hjemmeseier, borteseier eller uavgjort. Disse variablene vil være; form generelt, form på hjemme eller bortebane, resultat sist lagene møttes, resultat i ligaen forrige sesong og hjemmebanefordel.

3.2.1 Form generelt

For å predikere utfallet av en fotballkamp er det viktig å vurdere lagets nåværende form. Uttrykket «formlag» refererer til lag som er i god flyt, som betyr at de har tatt mange poeng i sine siste kamper. Dette gir en god indikasjon på hvilken form og selvtillit et lag går inn med i neste kamp. Vi har valgt å bruke resultatene fra de fem siste kampene som mål på et lags form (TromsøIL, 2023).

Begrunnelsen for valget av de fem siste kampene, er fordi dette gir en tilstrekkelig indikasjon på lagets prestasjoner den siste tiden. På den ene siden kan bruk av færre enn fem kamper gjøre at man ikke registrerer lagets reelle form, da enkelte resultater kan skyldes tilfeldigheter eller svak motstand. På den andre siden kan det å inkludere resultater fra flere enn fem

kamper resultere i et misvisende bilde av lagets form, ettersom prestasjoner kan variere betydelig. Vår tilnærming bekreftes av resultattabellen til Premier Leagues offisielle nettsted. Her kan man se at Premier League fremhever de fem siste kampene som en indikator på lags form, noe som underbygger vårt valg (Premier League, 2024).

3.2.2 Form på hjemme eller bortebane

Vi har valgt å inkludere formen til et lag basert på om laget spiller på hjemme- eller bortebane. Som tidligere nevnt har hjemmelag generelt en fordel, men effekten av fordelen kan variere mellom ulike lag. Derfor har vi valgt å spesifikt vurdere lagets prestasjoner i de tre siste kampene på respektive hjemme- eller bortebane. Denne tilnærmingen gir en bedre forståelse av lagets form, ved å vektlegge resultatene i forhold til hvor kampen finner sted. Dette gir en mer nøyaktig analyse av hvordan lagene presterer under forskjellig omstendigheter (Verma, 2024).

3.2.3 Resultat sist lagene møttes

Den tredje variabelen i modellen er resultatet fra sist lagene møttes. Når to lag møtes, kan både psykologiske og taktiske faktorer spille inn. Psykologiske faktorer kan være mentale overtak, mens taktiske faktorer kan være kombinasjon av spillestil. Historiske møter mellom lag kan avsløre mønster der et lag presterer bedre mot et annet lag, uavhengig av formen laget er i. Derfor tar modellen hensyn til tidligere møter mellom de respektive lagene, slik at man får med underliggende psykologiske og taktiske overtak.

3.2.4 Resultat i ligaen forrige sesong

En annen viktig variabel i modellen er lagets plassering på tabellen forrige sesong. Plasseringen reflekterer lagets prestasjon over tid, og gir en god indikasjon på hva man kan forvente av et lag i fremtiden. I Premier League er det de fire øverste lagene på tabellen som kvalifiserer seg til Champions League. Av de resterende lagene kvalifiserer de to neste seg til Europa League, og det syvende beste til Conference League (Brennan, 2023). Dette betyr at de syv øverste lagene på tabellen fra forrige sesong får konkurrere ute i Europa, noe som innebærer økte inntekter og attraktive vilkår for nye spilleranskaffelser. Dette gir bedre konkurranseevne, noe som fører til større sannsynlighet for topplassering i påfølgende sesong. Lag som derimot endte i bunnen av tabellen forrige sesong, vil starte neste sesong

med lav selvtillit og begrensede ressurser. Denne ulempen er noe som kan påvirke deres prestasjoner negativt (Peel & Thomas, 1992).

3.2.5 Hjemmebanefordel

Den siste variabelen inkludert i modellen er hjemmebanefordel. Som tidligere nevnt er det en stor fordel å spille kamper på hjemmebane. Ifølge en analyse av Li og Li (2023) viste resultatene fra Premier League sesongen 2022/2023 at 48,4% av kampene endte med hjemmeseier, og 26,3% resulterte i borteseier. Denne statistikken danner grunnlaget for hvordan hjemmebanefordelen vektes i modellen. Vi beregner forholdet mellom hjemme- og borteseier som følger:

$$\frac{48{,}42}{48{,}42+26{,}28} = 0{,}648 = 64{,}8\%$$

Formel 4 Beregning av hjemmebanefordel

Vekting av hjemmefordel er implementert i modellen som 65% sannsynlighet for seier til hjemmelaget og 35% sannsynlighet for seier til bortelaget (Li & Li, 2023).

3.3 Omkoding av datasett

Utgangspunktet for datasettet er illustrert i Tabell 1. Dette viser hvordan datasettet var utformet før omkodingen. I appendiks kan man se hvordan informasjonen er omkodet til en ny tabell. Denne tabellen er illustrert i Tabell 2, og er grunnlag for prediksjonsmodellen. Den inneholder de fem ulike variablene presentert i prosentverdier fordelt på hjemme- og bortelag.

	Div	Date	HomeTeam	AwayTeam	FTHG	FTAG	FTR	HTHG	HTAG	HTR	...	BbAv<2.5	BbAH	BbAHh	BbMxAHH	BbAvAHH	BbMxAHA	BbAvAHA	PSCH	PSCD	PSCA
0	E0	10/08/2018	Man United	Leicester	2	1	H	1	0	H	...	1.79	17	-0.75	1.75	1.70	2.29	2.21	1.55	4.07	7.69
1	E0	11/08/2018	Bournemouth	Cardiff	2	0	H	1	0	H	...	1.83	20	-0.75	2.20	2.13	1.80	1.75	1.88	3.61	4.70
2	E0	11/08/2018	Fulham	Crystal Palace	0	2	A	0	1	A	...	1.87	22	-0.25	2.18	2.11	1.81	1.77	2.62	3.38	2.90
3	E0	11/08/2018	Huddersfield	Chelsea	0	3	A	0	2	A	...	1.84	23	1.00	1.84	1.80	2.13	2.06	7.24	3.95	1.58
4	E0	11/08/2018	Newcastle	Tottenham	1	2	A	1	2	A	...	1.81	20	0.25	2.20	2.12	1.80	1.76	4.74	3.53	1.89

Tabell 1 Utgangspunktet for datasettet

	HomeFormPercent	AwayFormPercent	HomeFormPercentLast3	AwayFormPercentLast3	Intern_home_percent	Intern_away_percent	Result_last_season_H	Result_last_season_A	Home_team_advantage	Away_team_disadvantage
24	25.00	75.00	100.0	0.0	20.0	80.0	25.00	75.00	65	35
25	14.29	85.71	25.0	75.0	50.0	50.0	4.76	95.24	65	35
23	66.67	33.33	100.0	0.0	100.0	0.0	73.91	26.09	65	35
21	60.00	40.00	75.0	25.0	80.0	20.0	40.91	59.09	65	35
20	50.00	50.00	50.0	50.0	100.0	0.0	65.22	34.78	65	35

Tabell 2 Datasettet etter omkoding

Videre har vi valgt å utelukke de 20 første radene i datasettet. Disse radene inneholder informasjon om de første kampene som spilles i sesongen. Både variabelen *Form generelt* og *Form på hjemme- eller bortebane* baserer seg på resultatene fra siste kamper spilt. Dette innebærer at første serierunde, som er de første 10 kampene i datasettet, ikke har noen tidligere kamper å basere formen på. Det er når alle lag har spilt en hjemmekamp og en bortekamp, det vil si to serierunder, at formvariablene har et grunnlag. Ved å utelukke de to første serierundene vil vi derfor unngå NaN verdier, som vist i Tabell 3. Derfor har vi valgt å utelukke de to første kampene til hvert lag, det vil si de første 20 radene av datasettet.

	HomeTeam	AwayTeam	HomeFormLast3	AwayFormLast3
0	Man United	Leicester	NaN	NaN
1	Bournemouth	Cardiff	NaN	NaN
2	Fulham	Crystal Palace	NaN	NaN
3	Huddersfield	Chelsea	NaN	NaN
4	Newcastle	Tottenham	NaN	NaN
5	Watford	Brighton	NaN	NaN
6	Wolves	Everton	NaN	NaN
7	Arsenal	Man City	NaN	NaN
8	Liverpool	West Ham	NaN	NaN
9	Southampton	Burnley	NaN	NaN
10	Cardiff	Newcastle	NaN	NaN
11	Chelsea	Arsenal	NaN	NaN
12	Everton	Southampton	NaN	NaN
13	Leicester	Wolves	NaN	NaN
14	Tottenham	Fulham	NaN	NaN
15	West Ham	Bournemouth	NaN	NaN
16	Brighton	Man United	NaN	NaN
17	Burnley	Watford	NaN	NaN
18	Man City	Huddersfield	NaN	NaN
19	Crystal Palace	Liverpool	NaN	NaN

Tabell 3 Illustrering av variabelen «Form på hjemme eller bortebane»

3.4 Maskinlæring

Med maskinlæring menes maskinens evne til å lære basert på erfaring. Ved å tilføre data kan maskinen forbedre sin presisjon og yteevne. Maskinens evne til å lære vil derfor sees i sammenheng med å estimere avhengigheter ut ifra data. Fra dataen tilpasses en algoritme for å løse en gitt oppgave, uten direkte å bli programmert (Naqa & Murphy, 2015).

Tilpasningsprosessen, der algoritmen forbedres, kalles læringsdelen. I denne prosessen blir maskinen gitt et trengingssett. Av dataen konfigureres det en algoritme for å oppnå ønsket resultat basert på prøving og feiling. Gjennom trening blir dataen generalisert for å produsere resultat for ny og usett data (Naqa & Murphy, 2015).

3.5 Logistisk regresjon

Regresjonsanalyse anvendes for å etablere et forhold mellom mulige forklaringsvariabler og en responsvariabel (Thoresen, 2017). Den enkleste formen for regresjonsanalyse er lineær regresjon, som benyttes når responsvariabelen er kontinuerlig. Når vi skal predikere odds på Premier League kamper er det flere uavhengige variabler som påvirker den avhengige variabelen. Derfor bruker vi logistisk regresjon for å estimere sannsynligheten for spesifikke utfall. Klassisk logistisk regresjon er dikotom, det vil si at den avhengige variabelen har to mulige utfall. I vår prediksjonsmodell er det tre ulike utfall – seier, uavgjort, eller tap – noe som gjør at binær logistisk regresjon ikke kan benyttes. Vi bruker derfor multinomisk logistisk regresjon, fordi den avhengige variabelen ikke er begrenset til to verdier (IBM, 2024).

3.6 Fremgangsmetode

Variablene vi tar for oss vil gjennom maskinlæring bli vektet for å oppnå mest mulig presist resultat. For å gjøre dette blir modellen delt opp i et treningssett og et testsett. Treningssettet brukes til å trene opp modellen. Det vil si at modellen lærer om forholdet mellom de uavhengige variablene og den avhengige variabelen ut ifra hvor ofte den forutsier det riktige utfallet. Testsettet brukes til å evaluere hvor godt modellen kan predikere utfall på data den ikke har sett før. I vår modell har vi valgt å bruke 80% av dataen til treningssettet, mens de resterende 20% brukes til testsettet. Grunnen til at man splitter datasettet på denne måten, er for å unngå overtilpasning og kunne gi en realistisk evaluering av modellens prediktive kapasitet (Marthi, 2020).

Under prediksjonen estimerer modellen sannsynligheter for hver av de tre ulike utfallene i testsettet. Hver rad representerer en kamp, mens hver kolonne representerer sannsynligheten for de tre forskjellige utfallene som modellen har beregnet.

X-verdiene i den logistiske regresjonen er forklaringsvariablene. I denne modellen er det 5 uavhengige variabler i en dataframe som heter PM. Disse variablene er valgt som faktorer som kan påvirke utfallet av en fotballkamp. Y-verdien i modellen er den avhengige variabelen, med en verdi for faktisk resultat på fotballkampene. Denne verdien er uttrykt i en dataframe kalt FTR. Av dette skal vi predikere odds.

3.7 Anvendelse av modellen i praksis

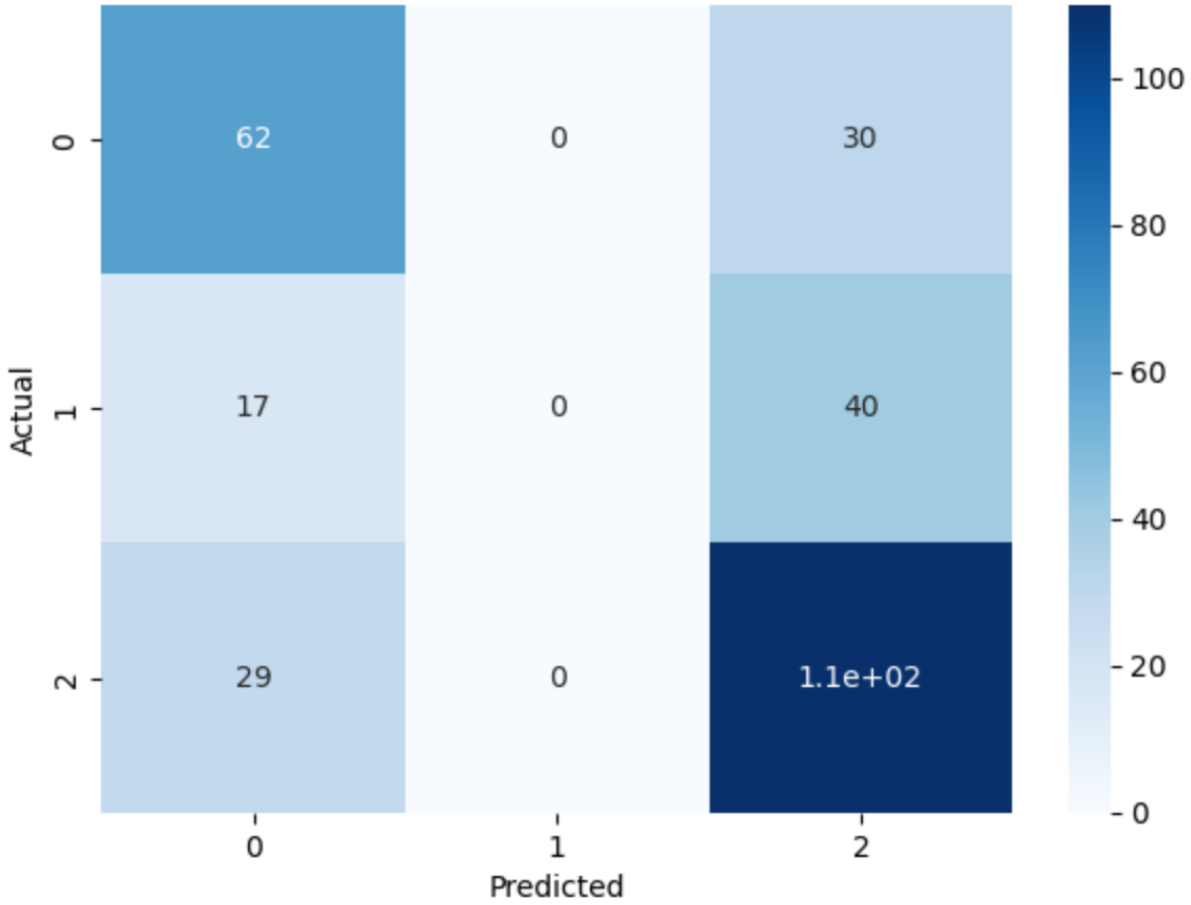
Formålet med prediksjonsmodellen er å fungere som et beslutningsverktøy for betting av Premier League kamper. For å estimere modellens potensielle lønnsomhet i praksis, kan man gjennomføre en simulering av betting basert på modellens prognoser. Dette innebærer bruk av en for-løkke som systematisk itererer seg gjennom alle kampene i både trenings- og testdatasettet. I hver iterasjon foretas det en sammenligning mellom modellens predikerte utfall og det faktiske kampresultatet. Avhengig av om prediksjonen stemmer eller ikke, vil simuleringen enten resultere i tap eller gevinst. Den potensielle gevinsten er lik innsatsen multiplisert med oddsen tilbudt av BET365 minus innsatsen. Gjennom denne metoden kan vi kvantifisere modellens evne til å generere en økonomisk fortjeneste over tid.

4. Resultat og analyse

4.1 Modellens presisjon

4.1.1 Presisjon på treningssett

Ut ifra resultatene på treningssettet kom man frem til at modellen har en treffsikkerhet på 59,7%, det vil si at modellen beregner riktig utfall i omtrentlig 6 av 10 tilfeller. Fordelingen av prediksjonene er illustrert i Figur 1. Figuren viser fordelingen av predikert verdi sett i sammenheng med faktiske verdier. Ut ifra dette kan man se at hyppigheten av predikert hjemmeseier og faktisk hjemmeseier dominerer. Videre predikerer modellen flere tilfeller av borteseier, totalt 108 ganger (29 + 17 + 62), hvor kun 62 av disse tilfellene stemte med borteseier resultatet.

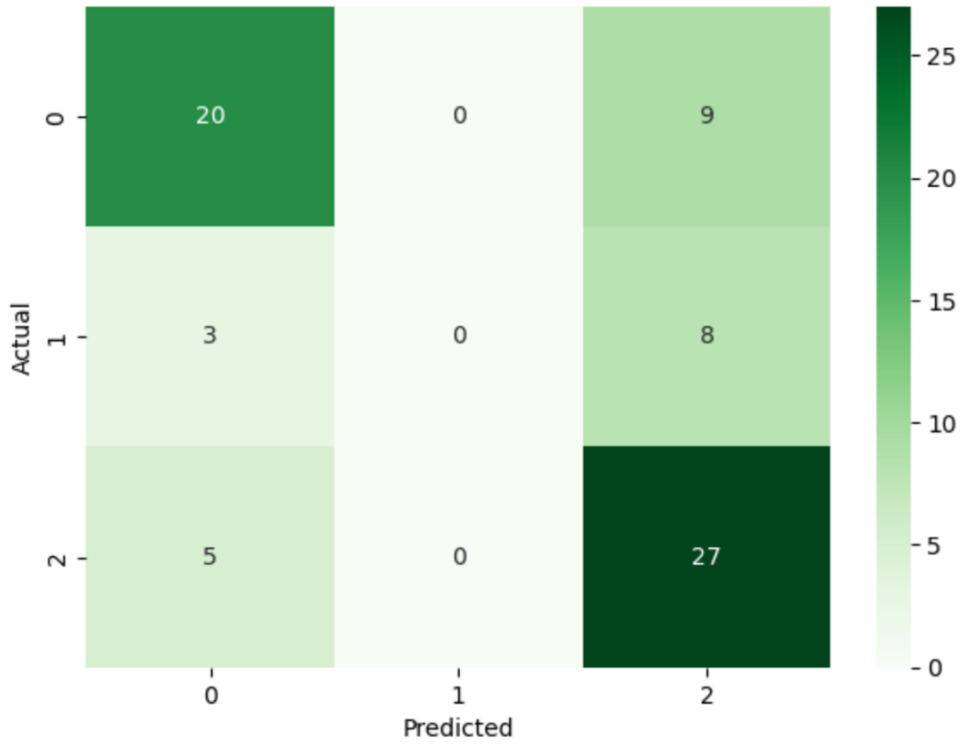


Figur 1 Presisjon på treningssettet illustrert i et Heatmap

4.1.2 Presisjon på testsett

I Figur 2 er fordelingen av testresultater illustrert. Av resultatene kom vi frem til at modellen har en treffsikkerhet 62,5%. Resultatene fordeler seg utover hjemme- og borteseier. Hjemmeseier er predikert i 44 tilfeller, hvor 27 av disse faktisk stemte. Forventet borteseiere var 28 ganger, hvor 20 av kampene faktisk endte med dette. Antall faktisk uavgjort var på 11, men modellen har i ingen tilfeller predikert dette. Etter å ha trent opp modellen på treningssettet kan man se at modellen har en *bedre* treffsikkerhet på testsettet enn på treningssettet. Selv om modellen har bedre presisjon på testsettet, har den likevel ikke prestert bedre i prediksjon av uavgjort.

Treffsikkerheten på 62,5% gjør det rimelig å anta at modellen har en generelt god evne til å generalisere dataen. Presisjonen er forholdsvis høy, samtidig som verdien ikke er *for* høy til å tro at dataen overtilpasses. Derfor kan det hevdes at modellen tenderer til å ikke overtilpasse treningsdataen og at resultatene er pålitelige.



Figur 2 Fordeling av presisjon på testsettet illustrert i et Heatmap

Resultatene på treningssettet og testsettet ga henholdsvis en presisjon på 59,7% og 62,5%. Den relativt tilsvarende nøyaktigheten mellom trening- og testsettet er et godt grunnlag til å anta at modellen er robust. At verdiene ligger forholdsvis tett viser at modellen er mindre følsom for små variasjoner i variablene.

Andre modeller i litteraturen, slik som Hucaljuk og Rakipović (2011), har hatt en treffsikkerhet på omtrent 60%. Dette kan illustrere at valg av variabler til denne modellen har vært gode, til tross for at dette er en forenklet modell av en kompleks idrett. Dette styrker troen på at modellen kan være et verdifullt verktøy for å hjelpe bettere til å ta informerte valg (Indeed Editorial Team, 2023).

4.2 Prediksjonsresultater

I Tabell 4 kan man se predikert odds i sammenheng med oddsen fra BET365. Overordnet er predikerte odds forholdsvis nært BET365, med unntak av noen utstikkere. Eksempelvis i kamp 257 var borteseier predikert til en odds på 24,59, hvor BET365 hadde satt oddsen til 3,2. Dette viser et tilfelle med stort avvik i forhold til oddsen satt av BET365.

	0	Prediction	Odds_H	Odds_D	Odds_A	B365H	B365D	B365A
157	1	2	1.46	6.01	6.79	2.30	3.40	3.40
141	2	2	2.44	3.89	3.01	3.00	3.10	2.70
263	2	2	1.68	4.21	6.03	2.87	3.40	2.62
205	2	2	1.66	4.05	6.61	2.14	3.10	4.20
57	0	0	9.82	7.96	1.29	1.57	4.33	6.50
...	...	...	...	...	...	...	...	...
152	2	2	1.29	5.72	20.40	2.90	3.20	2.75
257	2	2	1.24	6.44	24.59	2.35	3.50	3.20
50	2	2	1.53	4.27	8.91	5.00	4.00	1.75
196	2	2	1.82	4.31	4.56	3.30	3.20	2.25
255	0	0	4.25	5.14	1.75	1.22	7.50	14.00

Tabell 4 Prediksjon av kampresultat og odds på testsettet

```
Gjennomsnittlig avvik Odds_H: -1.88
Gjennomsnittlig avvik Odds_D: -3.06
Gjennomsnittlig avvik Odds_A: -3.69
```

Figur 3 Sammenlikning av BET365 sine odds og predikerte odds

Bettingselskap inkluderer risiko og profitt i sin fastsettelse av odds, noe som gjør at vi forventer gjennomsnittlig lavere odds i våre resultater sammenliknet med BET365. Ved å undersøke gjennomsnittlig avvik av alle oddsene viser resultatene at hjemmeseier hadde et avvik på -1,88, uavgjort på -3,06 og borteseier på –3,69, slik som illustrert i Figur 3.

Resultatene forteller oss at gjennomsnittet for hjemmeseier, uavgjort og borteseier alle har et negativt avvik, slik som forventet. Borteseier har det største avviket, noe som kan forklares av hjemmefordelen. Ettersom at hjemmebanefordelen er sterkt vektlagt i vår modell, og modellen påtar seg mindre premie for risiko, er dette avviket forventet større enn ved hjemmeseier. Grunnen til dette er fordi modellen kan tendere til å favorisere hjemmelaget. Uavgjort har også ett forholdsvis stort avvik, noe som kan skyldes at variabelen hjemmebanefordel ikke nyanserer uavgjort. Dette kan være grunnen til at uavgjort ikke blir et potensielt utfall.

	precision	recall	f1-score	support
0	0.71	0.69	0.70	29
1	0.00	0.00	0.00	11
2	0.61	0.84	0.71	32
accuracy			0.65	72
macro avg	0.44	0.51	0.47	72
weighted avg	0.56	0.65	0.60	72

Tabell 5 Classificationsreport av testsettet

I Tabell 5 kan man se at presisjonen på hjemmeseier er på 61%, borteseier på 71% og uavgjort på 0. Prediksjonsresultatene har som nevnt ovenfor en treffsikkerhet på 62,5%, men bommer likevel på prediksjon av uavgjort i 11 av 11 tilfeller, i følge Tabell 5. I trening av modellen var det null tilfeller av predikert uavgjort. Det kan derfor tenkes at algoritmen for uavgjort ikke er godt nok trent for å predikere dette, modellen ikke har hatt tilstrekkelig med data i treningssettet, eller at våre variabler har svakheter ved seg som gjør at modellen ser bort ifra uavgjort i alle mulige tilfeller.

4.3 Modellens evne til å generere økonomisk fortjeneste

For å teste om modellen er lønnsom har vi valgt å sette inn 100 kroner på hver kamp i både trening- og test settet. Modellen blir testet for hva potensiell gevinst ville væt dersom man hadde bettet på kampresultatet modellen foreslår. I treningssettet fikk vi en tilnærmet gevinst på 42 030 kroner. Basert på testsettet og dens 72 kamper viste simuleringen en total gevinst på 12 501 kroner. Resultatene er lønnsomme i både trening og testsettet. Modellen vil altså generere profitt og vil kunne være et godt verktøy i bettingbeslutninger.

5. Videre arbeid og feilkilder

5.1 Utvidelse av variabler

Vår prediksjonsmodell er begrenset på grunn av tidsrammen og størrelsen på oppgaven. I videre arbeid med modellen kan flere variabler inkluderes for en mer presis modell. Dette vil imidlertid kreve krever mer omfattende datasett og større innsats for å samle inn og analysere disse dataene. Selv om det er utfordringer knyttet til det å innhente inn mer kompleks data, vil dette sannsynligvis føre til mer nøyaktige resultater, ettersom flere relevante faktorer ville blitt tatt i betraktning (Oliver Hopkins, 2023).

En variabel som kunne vært inkludert i modellen er skader og suspensjoner blant nøkkelspillere. For eksempel spiller en spiss en avgjørende rolle på mange lag, fordi den ofte er toppscorer og dermed en viktig bidragsyter for å vinne kamper. Skader, sykdom eller suspensjoner på slike nøkkelspillere kan gjøre det utfordrende å finne fullverdige erstattere, noe som kan ha en betydelig påvirkning på kampresultatet. Lag med store spillertropper og mange gode erstattere vil sannsynligvis bli mindre påvirket av slike fravær sammenlignet med lag som har mindre spillerstaller. Derfor vil informasjon om tilgjengelighet av nøkkelspillere og kvaliteten på potensielle erstattere være verdifull til en utvidet modell.

En annen faktor som kan implementeres ved utbedring av modellen er antall hviledager mellom kamper. Fotballen har blitt en pengemaskin der inntekter på TV-rettigheter er kommet opp i mange milliarder kroner (Pedersen, 2023). Jo flere kamper og turneringer som spilles, desto mer tjener alle parter, slik som klubbene, sponsorer, rettighetshavere og forbund. Det innebærer et tett kampprogram, spesielt for topplagene som spiller turneringer ute i Europa. Med opptil tre kamper på en uke, vil det være store forskjeller på antall hviledager mellom kamper for de ulike lagene. Samtidig vil et tett kampprogram øke sannsynligheten for skader og gjøre det vanskelig å opprettholde samme intensitet gjennom hele kampen. For å finne nøyaktig antall hviledager for hvert lag må data fra alle kamper i de engelske cupene (FA Cup, Community Shield) og de europeiske turneringene (Champions League, Europa League, Conference League) inkluderes, noe som krever tilgang til flere datasett og mer detaljerte analyser.

En siste variabel som kunne vært implementert ved en større og bedre modell er plassering på tabellen før kampen som spilles. Dette er en variabel som krever at tabellen oppdateres etter

hver kamp og hver serierunde, noe som gjør det komplisert å implementere i modellen vår. Selv om vi allerede har variabler i modellen som baserer seg på formen til et lag (resultat siste kampene), er plasseringen på tabellen et godt bilde på hvordan et lag ligger an i forhold til laget de skal spille mot. Selv om et lag som ligger på andreplass på tabellen har tapt de to siste kampene sine, er de fortsatt favoritter til å vinne over et lag som ligger på sisteplass hvis de ligger høyt oppe på tabellen selv. Derfor er plassering på tabellen i skrivende stund en viktig formvariabel som man burde implementere ved en utbedring av modellen.

5.2 Feilkilder

Valg av å sløyfe variabler kan være en feilkilde i modellen. I utbredelsen av modellen og valg av data er det valgt å ikke inkludere resultater fra de førte 20 kampene spilt i ligaen. Optimalt sett burde disse kampene vært inkludert, og basert seg på data fra kamper forrige sesong. Dette er data som er komplekst å samle inn og fordi vår modell er en forenklet modell, var dette en begrensning som ble gjort.

Andre feilkilder er eventuell overlapping av variabler. I testing av modellen er det ikke korrigert for hvorvidt variablene måler det samme. Ved senere bruk av modellen foreslår vi derfor å undersøke om noen av variablene måler de samme faktorene. Grunnen til at overlapp kan være relevant å undersøke, er fordi dette kan være grunnen til hvorfor uavgjort ikke predikeres i modellen. Dersom noen variabler måler det samme er det en risiko for at denne feilen vektlegges i større grad en andre faktorer.

6. Konklusjon

Konklusjonen på oppgaven er at regresjonsmodellen kan benyttes til å ta informerte beslutninger om bettingvalg. Resultatene viser at modellen har en treffsikkerhet på 62,5%, noe som kan bidra til å drive betting med større grad av kontroll.

Modellen er god til å predikere hjemmeseier og borteseier, men har likevel en svakhet i prediksjon av uavgjort. I prediksjon av hjemmeseier er det færre avvik, sammenliknet med uavgjort- verdiene. Modellen er mindre tilpasset til å predikere uavgjort-verdier, da verken treningssettet eller testsettet faktisk predikere dette utfallet. Likevel vil modellens treffsikkerhet på 62,5% være et godt utgangspunkt i vurderingen av mulige kampresultater.

I videre arbeid er det mulighet for å legge til flere variabler for å utvikle en mer presis modell. Ved å inkludere flere variabler som eksempelvis formen til enkeltspillere og antall hviledager, vil dette kunne fange opp kompleksiteten i idretten og gi enda bedre prediksjonsresultater.

Simuleringen av modellen gir en positiv profitt for både trening- og testsettet. Ved å investere 100kr per kamp, viste resultatene på testsettet en profitt på 12 501 fordelt utover 72 kamper. Avslutningsvis kan man derfor konkludere med at modellen, gjennom en hel sesong, kan bidra til å ta informerte valg som generer profitt.

Litteratur

Forrest, D. & Simmons, R., 2002. Outcome Uncertainty and Attendance Demand in Sport: The Case of English Soccer. *Journal of the Royal Statistical Society.*

Peel, D. & Thomas, D., 1992. The demand for football: Some evidence on outcome uncertainty. *Empirical Economics* .

Hoås, T., 2012. *Bonuskampene,* s.l.: Empirical Economics (.

Stefánsson, O. & Bradley, R., 2019. What Is Risk Aversion?. *The British Journal for the Philsosphy of Sience.*

Sestovic, D., 2017. Bookmaker Margins and Favourite-Longshot Bias in Football Prediction Markets.

Michael, S. A., Paton, D. & Williams, L. V., 2009. Do bookmakers possess superior skills to bettors in predicting outcomes?. *Journal of Economic Behavior & Organization.*

Newall, P. W. S., 2023. How bookies make your money. *Judgment and Decision Makin.*

Yüce, S. G. et al., 2021. Effects of Sports Betting Motivations on Sports Betting Addiction in a Turkish Sample. *International Journal of Mental Healt and Addiction.*

Jeannotsson, A. A. & Ingvarsson, U. S., 2023. *What factors motivate and influence the decision of bettors to place bets on sports?,* s.l.: Department of Business Administration.

Hing, N., Russel, A. M. T., Vitartas, P. & Lamont, M., 2016. Demographic, Behavioural and Normative Risk Factors for Gambling Problems Amongst Sports Bettors. *Journal of Gambling Studies.*

Henriksen, M. K. & Nilsen, S., 2020. *Gambling i Norge,* s.l.: Norges arktiske universitet.

Na, S., Su, Y. & Kunkel, T., 2019. Do not bet on your favourite football team: the influence of fan identity-based biases and sport context knowledge on game prediction accuracy. *European Sport Management Quarterly,.*

Naqa, I. E. & Murphy, M. J., 2015. What Is Machine Learning?. *Machine Learning in Radiation Oncology.*

Hucaljuk, J. & Rakipović, A., 2011. Predicting football scores using machine learning techniques. *Proceedings of the 34th International Convention MIPRO.*

Kaggle, 2024. *Kaggle.* [Online]

Tilgjenglig fra: https://www.kaggle.com/datasets/saife245/english-premier-league

[Funnet Januar 2024].

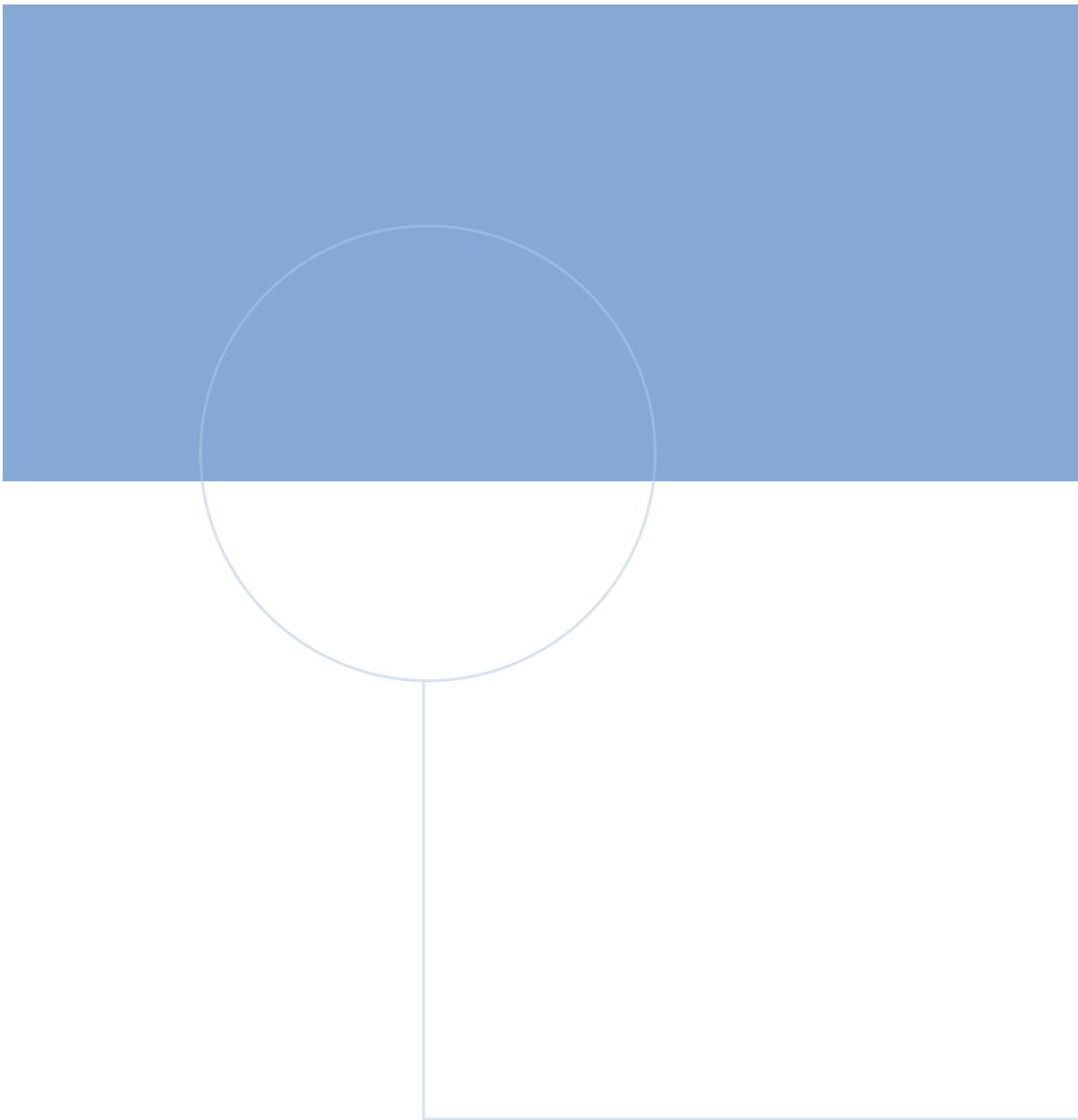
Morgan, S., 2018. *The Sun.* [Online]

Tilgjenglig fra: https://www.thesun.co.uk/sport/football/7065209/norwich-walls-liverpool-

floor-wimbledon-shower-football-dressing-room-tactics-revealed/
[Funnet Februar 2018].
Mood, C., 2010. Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It. *European Sociological Review.*
Jackson, J., 2023. *The Guardian.* [Internett]
Tilgjenglig fra: https://www.theguardian.com/football/2023/apr/02/police-investigating-attack-on-liverpool-team-bus-after-manchester-city-defeat
[Funnet Februar 2024].
Thoresen, M., 2017. Logistisk regresjon – anvendt og anvendelig. *Nor Legeforen* .
Li, M. & Li, J., 2023. Home Advantage in the English Premier League – Myth or Reality?. *Bruin Sports Analytics.*
Verma, R., 2024. *KhelNow.* [Internett]
Tilgjenglig fra: https://khelnow.com/football/2024-01-world-football-clubs-most-home-points-premier-league-2023
[Funnet Mars 2024].
Meyer, N., 2023. *Stack Exhange.* [Online]
Tilgjenglig fra: https://sports.stackexchange.com/questions/28857/how-often-has-a-team-beaten-another-team-from-the-same-league-four-times-in-a-se
[Funnet Februar 2024].
IBM, 2024. *Multinomial Logistic Regression.* [Internett]
Tilgjenglig fra: https://www.ibm.com/docs/en/spss-statistics/29.0.0?topic=regression-multinomial-logistic
[Funnet Mars 2024].
Oliver Hopkins, A. T. M. S., 2023. *Opta Analyst.* [Internett]
Tilgjenglig fra: https://theanalyst.com/eu/2023/10/arsenal-manchester-city-injuries-saka-rodri-martinelli-de-bruyne/
[Funnet Mars 2024].
Pedersen, M., 2023. *Nettavisen.* [Online]
Tilgjenglig fra: https://www.nettavisen.no/norsk-debatt/det-er-over-50-milliarder-grunner-til-at-pep-guardiola-sutrer-for-dove-orer/o/5-95-1270519
[Funnet Februar 2024].
Oddsenbonus, 2024. *Oddsenbonus.* [Internett]
Tilgjenglig fra: https://nhi.no/sykdommer/psykisk-helse/diverse/spilleavhengighet
[Funnet April 2024].

Complete Sports, 2024. *Comparing Popular Sports To Bet On Around The World.* [Internett]
Tilgjenglig fra: https://www.completesports.com/comparing-popular-sports-to-bet-on-around-the-world/
[Funnet April 2024].
Lindner, J., 2024. *Statistics About The Most Popular Sports In The World.* [Internett]
Tilgjenglig fra: https://gitnux.org/most-popular-sports-in-the-world/
[Funnet April 2024].
Zamain, A., 2023. *Why the Premier League is the best in the world.* [Internett]
Tilgjenglig fra: https://medium.com/@amoszamani6/why-the-premier-league-is-the-best-in-the-world-171b3836f4d6
[Funnet April 2024].
BET365, 2024. *How gambling works.* [Internett]
Tilgjenglig fra: https://responsiblegambling.bet365.com/responsible-gambling/how-gambling-works
[Funnet Mars 2024].
BET365, 2024. *About Us.* [Internett]
Tilgjenglig fra: https://help.bet365.com/en/about-us
[Funnet Mars 2024].
TromsøIL, 2023. *TIL.* [Online]
Tilgjenglig fra: https://www.til.no/nyheter/formlagene-barker-sammen
[Funnet Februar 2024].
Dobson, S. & Goddard, J., 2011. *The Economics of Football.* Bangor: Cambridge University Press.
Premier League, 2024. *Premier League.* [Internett]
Tilgjenglig fra: https://www.premierleague.com/tables
[Funnet April 2024].
Brennan, F., 2023. *European places in Premier League for 2023/24: How Champions League, Europa and Conference League are decided.* [Online]
Tilgjenglig fra: https://www.sportingnews.com/us/soccer/news/european-places-premier-league-2023-24-europa-league/qoru7fin1dbewxrtixheslcl
[Funnet April 2024].
Indeed Editorial Team, 2023. *What Is a Confusion Matrix? (Plus How To Calculate One.* [Online]

Tilgjenglig fra: https://www.indeed.com/career-advice/career-development/confusion-matrix
[Funnet April 2024].
Helsedirektoratet, 2022. *Pengespel og avhengnad.* [Internett]
Tilgjenglig fra: https://www.helsenorge.no/rus-og-avhengighet/spillavhengighet/
[Funnet April 2024].
Marthi, G., 2020. *Machine Learning Basics: Logistic Regression.* [Online]
Tilgjenglig fra: https://towardsdatascience.com/machine-learning-basics-logistic-regression-890ef5e3a272
[Funnet Mars 2024].



NTNU

Kunnskap for en bedre verden