Wine and Food Pairing Predictor

Ivan Eirik Nesbø, Oliver Hansen Kristiansen, Vegard Krieg

# **1: Problembeskrivelse**

**OMFANG**

Vi ønsker å lage en løsning for hvilken vin man skal kjøpe sammen med et godt måltid.

I dag løses dette problemet gjennom å spørre på vinmonopolet, og satse på at vedkommende har nok kunnskap til å gi en god anbefaling. Med denne løsningen får man gjerne en subjektiv tilbakemelding, og ikke en felles konsensus.

Vår løsning vil være tilgjengelig for alle og enhver, og lett tilgjengelig på nett. Du skal kun trenge å skrive inn hvilket måltid du tenker å lage, og få anbefalinger basert på en stor samling av vurderinger. Fordi løsningen er lettvint og tilgjengelig, vil det være en go-to ressurs for alle som liker god vin til mat.

## **METRIKKER**

For å evaluere hvor godt maskinlæringsmodellen fungerer bruker vi flere metrikker. I hovedsak bruker vi klassifikasjonsmetrikker fra Scikit-learn biblioteket, som accuracy, precision, recall og F1-score. Disse metrikkene er viktig å bruke for å være sikker på at det som predikeres faktisk stemmer overens med datasettet.

Vi har underveis i arbeidet lagt inn metrikker for å ha oversikt over hvor godt modellen fungerer. Her printer vi metrikkene som tidligere beskrevet, samt med eksempelinput av matrett.

# **2: DATA**

I arbeidet med løsningen har vi innhentet et stort dataset med 34934 rader. Dette datasettet har kolonnene;

* Wine\_type - string
* Wine\_category - string
* Food\_item - string
* Food\_category - string
* Cuisine - string
* Pairing\_quality - integer
* Quality\_label - string
* Description - string

Datasettet er henta fra kaggle.com, data kan og hentes fra resturanter, vinmonopol eller vineksperter og undersøkelser.

34934 datapunkter er vanligvis nok for supervised learning hvor det er anbefalt med å ha en modell med mellom 30 – 50 tusen datapunkter.

Innledningsvis brukte vi flere av kolonnene til å predikere utfallet, men etter å ha testet modellen, snevret vi det ned til kun food\_item og food\_category. Dette er fordi kolonner som cuisine bare skapte støy uten å gi modellen noe ytterligere informasjon. Vi filtrerte også ut alle pairing\_quality som var <= 4, slik at vi kun beholdt data med 4 eller 5 i pairing\_quality.

Kolonnene som predikeres er wine\_category og wine\_type. Det vises kun vintyper innenfor kategorien som predikeres til brukeren.

# **3: MODELLERING**

Modellen vi har valgt å bruke er RandomForestClassifier. Dette er fordi denne modellen håndterer kategoriske features og multi-class classification på en veldig god måte. Den håndterer også komplekse kombinasjoner uten problemer. Vi bruker OneHotEncoder på treningsdataene og Label Encoder på target.

Når vi deklarerte modellen, ble n\_estimators satt til 100. Vi forsøkte å sette denne høyere, men dette ga samme accuracy ved n=200 og n=300. Vi forsøkte også å bruke andre parametere, men fant ikke noe suksess med dette.

# **4: DEPLOYMENT**

Modellen settes i drift som en enkel web-applikasjon der brukeren velger en matrett som input og får anbefalt en vin. Dette gjøres ved bruk av Gradio, et Python-bibliotek som lar oss lage et interaktivt grensesnitt direkte fra koden. I applikasjonen møter brukeren en dropdown-meny hvor man kan velge en matrett. Modellen bruker deretter treningsdataene til å predikere både vin-kategori og vin-type. Resultatet presenteres tydelig som en rangering av de fem vinene som scorer høyest.

Applikasjonen kan kjøres lokalt eller deles via en offentlig URL generert av Gradio. Dette gjør det enkelt å demonstrere og teste løsningen med flere brukere.

For videreutvikling og skalering kan modellen integreres i et Spring Boot-miljø og kobles til en database med oppdaterte data om mat- og vinparinger. Databasen kan da kontinuerlig utvides med ferske data og nye trender innen mat og vin.

# **5: KI-DEKLARASJON**

I arbeidet med denne modellen har vi brukt kunstig intelligens til følgende:

* feilsøking
* retting av syntaks
* forståelse av datasettet
* forståelse av konsepter i maskinlæring

Ingen kode er klippet og limt, all kode er selvskrevet.

# **6: REFERANSER**

1. Khr, Alex. *Wine and Food Pairing EDA & Prediction (acc = 1.0)*. Kaggle, <https://www.kaggle.com/code/alexkhr/wine-and-food-pairing-eda-prediction-acc-1-0/notebook>.
2. El Husseini, W., & 1 collaborator. (2025, October). *Wine and food pairing dataset*. Kaggle. [https://www.kaggle.com/datasets/wafaaelhusseini/wine-and-food-pairing-dataset](https://www.kaggle.com/datasets/wafaaelhusseini/wine-and-food-pairing-dataset?utm_source=chatgpt.com)
3. Gradio Documentation,

<https://www.gradio.app/docs/gradio/>

1. Scikit-learn.org

<https://scikit-learn.org/stable/modules/ensemble.html#forest>