

# Projekt taxi prediction

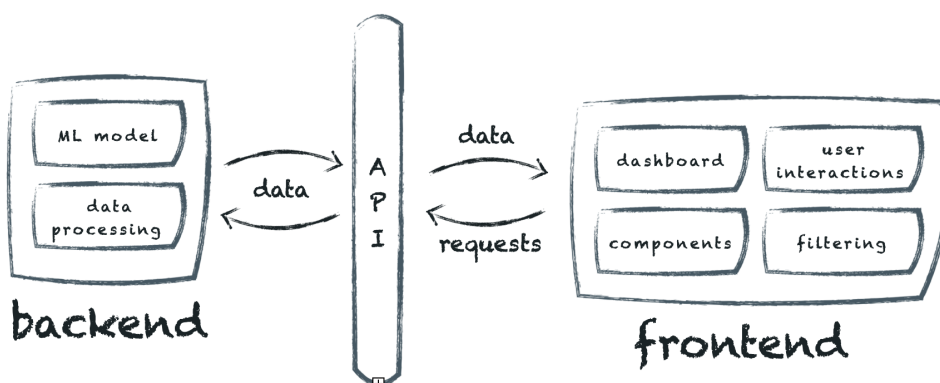


## Syfte

Projektet syftar till att implementera backend och frontend för att lösa ett verklighetsnära problem. Backend och frontend kommunicerar via ett API-lager för att göra komponenterna decoupled från varandra. Applikationen kommer servera en machine learning modell för att göra relevanta predictions.

## Bakgrund

Företaget Resekollen AB har tidigare byggt applikationer för att visualisera och planera resor med kollektivtrafiken. Nu vill de expandera till ett närliggande område, nämligen taxiresor. De vill kunna prediktera taxipriser och har anlitat dig som LIA studerande för att hjälpa dem med detta. Med dina kunskaper inom machine learning, OOP, API, backend och frontend och säkert andra fina keywords också kommer det gå galant. En tidigare machine learning engineer (MLE) har tillsammans med product owner (PO) skissat upp följande arkitektur och har byggt grundläggande kodstruktur som ni ska bygga vidare på. Dessvärre blev MLE headhuntad till ett annat bolag och valde att säga upp sig. Innan han lämnade så har han också skapat grundstruktur i koden som ni kommer få ta över.



VDn visade detta för sina investerare och fick in riskkapital från bland annat Yohanna. Yohanna som tidigare investerat i dataintensiva projekt förstår vikten av noggrannhet, då det oftast straffar sig av att ta genvägar. Därför har hon gett Börje som är PO mer befogenheter att göra det här projektet noggrant. Dessutom är det viktigt med fler funktionaliteter i applikationen som ger bättre kundupplevelse.

## Roller i organisationen

Person	Roll	Beskrivning
Yohanna	riskkapitalist	investerat i dataintensiva projekt, begränsad teknisk kompetens, bra businesskunskaper
Roager	VD	tuff förhandlare, duktig på business
Börje	product owner	begränsad teknisk kompetens, bra kontakt med både Yohanna och Jacobi
Feliz	frontendutvecklare	sitter för närvarande med annat projekt, men kan assistera med git och github
Yulia	BI-utvecklare	sitter ute hos kund

## Scenario

Yohanna har varit med i tidigare projekt som letts av studerande på yrkeshögskolan inom data engineering och varit mycket nöjd över resultaten. Därför bestämmer sig Yohanna tillsammans med Roager att ta in LIA till detta projekt också.

Teamet i Reskollen skissar hastigt en kravställning.

## Uppgift 0 - Uppvärmning

Det här projektet ska göras ensamt men samarbeta gärna med flera personer. Varje person skickar in egen githublänk till lärplattformen, men skriv vem/vilka ni jobbat tillsammans med. Skapa ett githubrepo som heter taxi-prediction-fullstack-<FÖRNAMN>-, ersätt med ditt namn.

Kopiera in filstrukturen från lektion a2\_packaging

[Datan hittar du här.](#)

Läs på om metadatan så du vet vad de olika kolumnerna betyder. Du kommer ha nytta av detta när du ska göra EDA, servea datan via APIet och göra ML modell.

## Uppgift 1 - EDA och data cleaning

Starta i en eller flera Jupyter notebooks och gör explorativa dataanalyser för att förstå datasetet och eventuellt rita en del grafer. Du har en del nullvärden i datasetet, fundera på lämpliga sätt att städa datan. Vissa värden kanske kan ersättas med 0? Andra värden kanske kan ersättas med något annat? En del rader kan man kasta helt.

Vill du ta enkla vägen så kan du kasta de flesta raderna med nulls. Se dock till att spara de raderna där label saknas - dessa kan med fördel användas som user inputs för att få en prediction av resepriset.

Se också till att ta bort eventuella outliers, exempelvis i priset då outliers enkelt kan påverka modellens noggrannhet. Ni kan exportera datan efter ni processat den så kan ni använda den rensade datan i applikationen istället för rådatan.

## Uppgift 2 - ML modell

Ta fram en ML modell för att prediktera taxipriser. Gå igenom data science workflow, med att testa olika modeller och evaluera. Välj därefter en modell och träna på all data. Exportera därefter datan med hjälp av `joblib`.

## Uppgift 3 - Backend och API-lager

Skapa en backend och ett API-lager i FastAPI som ska servea datan. Det innebär att det ska finnas olika endpoints för att kunna läsa delar av datan. Det ska också finnas möjlighet att göra prediktion genom att låta användaren mata in olika parametrar.

## Uppgift 4 - Bygg en frontend

Bygg en snygg frontend och konsumera API:et.

## Uppgift 5 - Videopitch

Spela in dig och själv och skärmen - ex mha teams, obs eller liknande verktyg. Du ska visa upp applikationen och presentera övergripande hur din kod fungerar. Videon ska vara 5-10 min lång och skickas in till lärplattformen.

## Bedömning

Projektet bedöms individuellt, men jobba gärna tillsammans med andra. Dock kom ihåg att skriva ned vem/vilka ni jobbat tillsammans med.

## Godkänt

- gjort uppgifterna korrekt
- gjort flera relevanta committs
- kunna kommunicera med korrekt terminologi
- kunna kommunicera sitt projekt på ett förståeligt sätt