

Uppdragsbeskrivning

Ledningen på E-handelsgiganten KokoBananas gick för ett antal år sedan helt över till digital marknadsföring & försäljning. Det visade sig vara ett klokt val. Med tiden har dock, i takt med att konkurrenternas digitala kompetens höjts, även den digitala konkurrensen hårdnat. Deras bransch, den kosmetiska & hälsoinriktade, är särskilt utsatt. Ledningen har nu kommit till insikten om att nyckeln till framgång inom E-handel är besökernas användarupplevelse!

Ett problem de märkt är att alltför många besökare på deras hemsida inte konverterar, dvs inte köper något. I syfte att förstå vad en bra användarupplevelse är, behöver företaget således först förstå vilka kunder som faktiskt konverterar och köper.

Du har fått i uppgift att inledningsvis utveckla en 'AI-modell' som kan förutse om en kundsession kommer att resultera i ett köp eller inte, i realtid.

Till din hjälp har du historisk data på kundsessioner innehållande förutsättningar som gäller för varje session, samt huruvida respektive session leder till ett köp eller ej. Ambitionen och förhoppningen är att resultatet på denna analys ska kunna appliceras på framtida kunder, i realtid – för att snabbt kunna identifiera de kunder som faktiskt beställer produkter.

När du gjort dina analyser ska du sedan presentera dina slutsatser för företagsledningen – på ett begripligt sätt. Både styrkor & svagheter i dina prediktioner behöver lyftas, så att ledningen & företaget kan ta korrekta beslut baserade på modellens genererade beslutstöd.

Kommentarer från Ali

Tekniskt sett blir detta alltså ett binärt klassificeringsproblem. Du ska utföra den angivna uppgiften och resultatet av detta arbete kommer, vid ett senare skede, att användas för att förbättra samtliga kunders användarupplevelse.

Du kan läsa om själva kolumnerna i datasetet [här](#) & [här](#). Det är dock datasetet i ITHS-distans som ni skall använda!

Presentationen för ledningen är en användarvänlig och läsbar redovisningen av dina upptäckter och slutsatser – och skrivs ned i slutet av din notebook. Hela projektet inklusive 'presentationen för ledningen' sker i en sammahållen Jupyter Notebook.

Du kommer genomgående även bedömas på hur begriplig din kod är samt hur bra du kommenterar och motiverar dina beslut. Det ska vara enkelt för mig att följa dina tankegångar.

Beslutskriterier:

- **G (Godkänt):** För att få godkänt betyg bör studenten visa förståelse och färdighet i att tillämpa maskininlärningsmetodik på ett, i stort, korrekt sätt. Studenten förväntas använda tekniker de har lärt sig i kursen för att utveckla och utvärdera ML-modeller.
- **VG (Väl Godkänt):** För att få väl godkänt krävs det att studenten, utöver kraven för godkänt, visar mycket god förståelse för de maskininlärningsmetoder vi lärt oss i kursen. Studenten ska även visa mycket god förmåga i att både utveckla och evaluera ML-modeller – samt presentera de erhållna resultaten på ett begripligt sätt.

Generella bedömningspunkter:

Studenten väljer själv tillvägagångssättet under det självständiga arbetet. Således krävs inte att studenten använder samtliga metoder & verktyg vi lärt oss i kursen, utan enbart de studenten själv anser är nödvändiga för att lösa uppgiften.

Dvs, trots att allt inte strikt är obligatoriskt sker bedömning generellt på följade punkter:

1. Dataförståelse och förberedelse:

- Utforska och förstå det givna datasetet.
- Utföra en explorativ dataanalys för att identifiera mönster och samband i data.
- Identifiera och behandla eventuella brister eller inkonsekvenser i data (t.ex. saknade värden, outliers, onormala värden).
- Utföra eventuell feature engineering, feature scaling och dimensionsreducering.
- Undvika dataleakage.

2. Modellval och utvärdering:

- Utval av lämpliga inläringssalgoritmer baserat på problemets art.
- Tillämpa valideringsmetoder för att utvärdera prestanda.
- Jämför och analysera resultaten för att välja den mest effektiva modellen.
- Utföra lämplig error-analys.

3. Optimering och hyperparameterjustering:

- Utför hyperparameterjustering för att optimera modellernas prestanda.
- Använd tekniker som exempelvis grid search för att söka efter de optimala hyperparametrarna för givna modeller.

4. Insiktsutvinning och rapportering:

- Extrahera insikter från den slutliga modellen för att förstå dess praktiska användbarhet.
- Presentera resultaten för ledningen på ett begripligt sätt i syfte att underlätta beslutsfattandet.

Den slutgiltiga betyget baseras på en helhetsbedömning

Inlämningsinstruktioner:

Inlämning av notebook innehållande hela arbetet sker på ITHS-portalen, på samma plats ni laddade ner dessa instruktioner ifrån.

Namnge notebook:en enligt *namn_efternamn_ml_projekt.ipynb*.

OBS! För min skull, spara all output i notebook:en när ni lämnar in den - så att jag idealt ej behöver köra er kod för att se er output.