



Machine Learning

2024

Inlämning 2

Machine Learning Techniques

Analyzing E-commerce Data

Översikt

Denna uppgift innebär en praktisk övning med Amazon e-handelsdataset. Huvudfokus kommer att ligga på att förstå och analysera resultat från en valfri maskininlärningsmodell. Uppgiften är centrerad kring ett nyckelaspekt av e-handel: kundbeteende och köpmönster.

Syfte

- **Huvudmål:** Förutspå (predict) kunders utgifter (spending) ('Amount') baserat på tillgängliga funktioner (features) som 'Category', 'Size' och 'Quantity'.
- **Fokus på Analys:** Snarare än att bygga flera komplexa modeller, ligger tonvikten på djup förståelse och analys av modellens resultat.

Uppgifter

1. Modellval:

- Välj en enda maskininlärningsmodell som du anser lämplig för regressionsuppgifter. Detta kan vara Linear Regression, Decision Trees eller vilken modell du än är bekväm med.
- Motivera ditt val av modell med avseende på dess lämplighet för uppgiften och enkelhet i tolkningen.

2. Dataförberedelse:

- Genomför nödvändiga förberedande (Data preprocessing) steg såsom hantering av saknade värden, kodning av kategoriska variabler och delning av datasetet i tränings- och testuppsättningar.

3. Modellträning och Förutsägelse:

- Träna din valda modell på datasetet.
- Använd modellen för att göra förutsägelser på testsatsen.

4. Fördjupad Analys av Resultaten:

- Analysera modellens prestanda med hjälp av lämpliga mätvärden (t.ex. Mean Squared Error, R-squared).
- Tolkning av resultaten: Vad säger modellens förutsägelser oss om kundbeteendet? Vilka funktioner påverkar mest kundens utgifter (spending)?
- Diskutera eventuella potentiella fördomar eller begränsningar i din modell och hur de kan påverka resultaten.

5. Reflektion över Modellvalet:

- Reflektera över ditt val av modell: Hur underlättade eller hindrade det din analys? Skulle en annan modell eventuellt ge bättre insikter?

Leveranser

- **Rapport:** Ett omfattande dokument som täcker din metodik, analys av modellens resultat, insikter om kundbeteende baserat på modellens förutsägelser och reflektioner över valet av modell.
- **Presentation:** En koncis presentation som sammanfattar viktigaste fynden, modellens prestanda och insikter som härleds från analysen.

Dokumentet kan vara kodfilen (Jupyter notebook) som innehåller alla kommenterna och charts & graphs eller en separat text-fil.