

## **Izvršni sažetak**

### **Uvod**

U 2019. godini, New York je zabilježio preko 103 milijuna vožnji taksijem, što predstavlja bogat skup podataka koji odražava složene obrasce urbane mobilnosti. S obzirom na to da prometni sektor doprinosi gospodarstvu grada s otprilike 15 milijardi dolara godišnje, optimizacija predviđanja cijena vožnji ključna je za poboljšanje operativne učinkovitosti i zadovoljstva korisnika. Ovaj se projekt fokusira na predviđanje iznosa cijena taksi vožnji pomoću tehnika strojnog i dubokog učenja. Projekt je usklađen s inicijativom pametnog grada jer koristi podatke u stvarnom vremenu za održiva i inteligentna rješenja mobilnosti.

### **Definicija problema**

Precizno predviđanje cijena taksi vožnji i dalje predstavlja izazov u dinamičnom okruženju poput New Yorka, gdje karakteristike vožnje kao što su udaljenost, promet i doba dana neprestano variraju. Netočne predikcije mogu dovesti do nezadovoljstva korisnika i neučinkovitosti u određivanju cijena. Cilj ovog projekta je riješiti te probleme razvojem modela koji koriste povijesne i kontekstualne podatke za precizniju procjenu cijena vožnji.

### **Ciljevi**

1. Predvidjeti iznose cijena taksi vožnji koristeći povijesne podatke o vožnjama.
2. Klasificirati vožnje u cjenovne kategorije radi pojednostavljene analize.
3. Ocijeniti modele strojnog i dubokog učenja za procjenu cijena vožnji.
4. Analizirati značajke koje utječu na varijabilnost cijena.
5. Potvrditi (validirati) performanse modela pomoću statističkih metrika.
6. Istražiti klasteriranje za segmentaciju ponašanja putnika.
7. Ponuditi praktične preporuke za primjenu u stvarnom svijetu.

### **Metodologija**

Studija primjenjuje modele k-Nearest Neighbors, Linear Regression, Decision Trees, SVR, Random Forest, Gradient Boosting i modele dubokog učenja za predviđanje cijena vožnji. Predobrada podataka uključivala je inženjering značajki, poput vremenskih i prostornih varijabli. Također je provedeno klasteriranje pomoću algoritama K-Means i DBSCAN kako bi se identificirali obrasci korištenja.

### **Ključni nalazi**

- Linearna regresija postigla je RMSE od 2,93 i  $R^2$  od 0,958, pokazujući visoku točnost.
- Modeli Random Forest i Gradient Boosting pokazali su robusne performanse, pri čemu je Random Forest bio neznatno bolji u regresijskim zadacima.
- Model dubokog učenja imao je najniži MAE (1,48), bio je izvrstan u predviđanju niskih cijena, ali je blago podcjenjivao visoke cijene.

- Klasifikacijski modeli poput logističke regresije i modela Random Forest postigli su točnost iznad 94% i F1-rezultate iznad 0,90.
  - Klasteriranje pomoću K-Means ( $k=3$ ) pružilo je optimalno razdvajanje prema siluetnim ocjenama (silhouette scores), ističući klastere obrazaca vožnji.
- 

## **Preporuke**

- Uključiti podatke o vremenu, prometu i događajima kako bi se poboljšala točnost predviđanja.
- Primijeniti tehnike balansiranja podataka kako bi se riješio problem neravnoteže u cjenovnim kategorijama.
- Razmotriti arhitekture poput LSTM-a ili Transformera za bilježenje vremenske dinamike.
- Razviti nadzornu ploču (dashboard) u stvarnom vremenu za predviđanje cijena kao pomoć vozačima i putnicima.
- Koristiti geoprostorno klasteriranje za precizniju analizu temeljenu na lokaciji.

## **Zaključak**

Ovaj projekt pokazuje da kombinacija tradicionalnog strojnog učenja, dubokog učenja i klasteriranja može učinkovito modelirati dinamiku cijena taksi vožnji. Ovi uvidi podupiru odluke temeljene na podacima u planiranju urbane mobilnosti te mogu poboljšati određivanje cijena, raspoređivanje vozila (dispečing) i korisničko iskustvo. Sve u svemu, modeli su imali najbolje performanse za kratke vožnje i niže cjenovne raspone, gdje je koncentrirana većina taksi aktivnosti.