

# Sažetak

## Uvod

U 2019. godini, New York City je zabilježio preko 103 milijuna putovanja taksijem, što predstavlja bogat skup podataka koji odražava složene obrasce urbane mobilnosti. S obzirom na to da prometni sektor godišnje doprinosi gradskom gospodarstvu s otprilike 15 milijardi dolara, optimizacija predviđanja cijena prijevoza ključna je za poboljšanje operativne učinkovitosti i zadovoljstva korisnika. Ovaj projekt usmjeren je na predviđanje iznosa cijena taksija korištenjem tehnika strojnog učenja i dubokog učenja. Usklađen je s inicijativom pametnog grada korištenjem podataka u stvarnom vremenu za održiva i inteligentna rješenja za mobilnost.

## Izjava o problemu

Točno predviđanje cijena taksija ostaje izazov u dinamičnom okruženju poput New Yorka, gdje karakteristike putovanja poput udaljenosti, prometa i doba dana stalno variraju. Netočna predviđanja mogu dovesti do nezadovoljstva kupaca i neučinkovitosti u određivanju cijena. Cilj ovog projekta je riješiti te probleme razvojem modela koji koriste povijesne i kontekstualne podatke za precizniju procjenu cijena.

## Ciljevi

1. Predvidite iznose taksi prijevoza koristeći povijesne podatke o putovanju.
2. Klasificirajte putovanja u kategorije cijena radi pojednostavljene analize.
3. Procijenite modele strojnog učenja i dubokog učenja za procjenu cijene prijevoza.
4. Analizirajte značajke koje utječu na varijabilnost cijena.
5. Validirajte performanse modela pomoću statističkih metrika.
6. Istražite klasteriranje za segmentaciju ponašanja putnika.
7. Ponudite praktične preporuke za primjenu u stvarnom svijetu.

## Metodologija

Studija primjenjuje modele k-najbližih susjeda, linearne regresije, stabala odlučivanja, SVR, slučajne šume, gradijentnog pojačavanja i dubokog učenja za predviđanje cijena. Prethodna obrada podataka uključivala je inženjering značajki kao što su vremenske i prostorne varijable. Također je provedeno klasteriranje s K-means i DBSCAN-om 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.
- Random Forest i Gradient Boosting pokazali su robusne performanse, s Random Forestom koji je neznatno nadmašio rezultate u regresijskim zadacima.
- Model dubokog učenja imao je najnižu MAE (1,48), izvrsno predviđajući niske cijene, ali neznatno podcjenjujući visoke cijene.
- Klasifikacijski modeli poput logističke regresije i slučajne šume postigli su točnost preko 94% i F1 rezultate iznad 0,90.
- K-Means grupiranje ( $k=3$ ) osiguralo je optimalno odvajanje po rezultatima siluete, ističući klastere uzoraka putovanja.

## Preporuke

- Uključite podatke o vremenu, prometu i događajima kako biste poboljšali točnost predviđanja.
- Primijenite tehnike uravnoteženja podataka kako biste riješili neravnoteže u klasama cijena.
- Razmotrite LSTM ili Transformer arhitekture za snimanje vremenske dinamike.
- Razviti nadzornu ploču u stvarnom vremenu za predviđanje cijena kako bi se pomoglo vozačima i putnicima.
- Koristite geoprostorno grupiranje za poboljšanje analize temeljene na lokaciji.

## Zaključak

Ovaj projekt pokazuje da kombiniranje tradicionalnog strojnog učenja, dubokog učenja i klasteriranja može učinkovito modelirati dinamiku cijena taksija. Ovi uvidi podržavaju odluke utemeljene na podacima u planiranju urbane mobilnosti i mogu poboljšati cijene, otpremu i korisničko iskustvo. Općenito, modeli su se najbolje pokazali za kratka putovanja i niže raspone cijena, gdje je koncentrirana većina aktivnosti taksija.