

Sistemi za podršku odlučivanju i kreiranje izvještaja

Analiza i predviđanje kašnjenja letova

- Opis problema

Danas se suočavamo s ozbiljnim izazovom povećanja kašnjenja i otkazivanja letova. Ova neprijatnost ne samo da troši resurse u transportu, već i značajno utiče na planove putnika, dovodeći do povećanja nezadovoljstva i broja pritužbi. Nezadovoljstvo putnika i nedostatak povjerenja u aviokompanije značajno narušavaju korporativni ugled, što dalje utiče na vjernost putnika. Problematika koja je odabrana za analizu obuhvata detaljne informacije o dolascima i kašnjenjima letova na aerodromima u SAD-u, pri čemu su podaci kategorisani po prevoznicima. Razvijeni sistem će iskoristiti ove podatke o prethodnim letovima kako bi predvidio moguća kašnjenja letova. Ova analiza ima potencijal da pomogne putnicima u planiranju svojih putovanja, dok avio-kompanijama pruža uvid u oblasti koje se mogu unaprijediti radi poboljšanja ukupne operativne efikasnosti.

- Skup podataka

Skup podataka koji će se koristiti u projektu se može naći na Kaggle-u i pruža uvid u performanse avio-prevoznika na različitim aerodromima u SAD tokom perioda od avgusta 2013. do avgusta 2023. godine, fokusirajući se na dolaske letova i kašnjenja. Skup podataka je struktuiran u tabelarnom formatu, pri čemu redovi predstavljaju jedinstvene kombinacije godine, mjeseca, prevoznika i aerodroma. Svaki red sadrži informacije o različitim metrikama, uključujući broj letova, broj kašnjenja, broj otkazivanja i preusmjerenja, kao i raspodjelu kašnjenja prema različitim faktorima. Kolone pružaju specifične detalje poput kodova i imena prevoznika, kodova i imena aerodroma, kao i broja kašnjenja koja se pripisuju prevozniku, vremenskim uslovima, Nacionalnom vazduhoplovnom sistemu (NAS), sigurnosti i kašnjenjima prilikom dolaska aviona. Struktuirani format omogućava korisnicima lako pretraživanje, analiziranje i vizualizaciju podataka radi izvođenja smislenih zaključaka.

- year: Godina podataka.
- month: Mjesec podataka.
- carrier: Kod prevoznika.
- carrier_name: Ime prevoznika.
- airport: Kod aerodroma.
- airport_name: Ime aerodroma.
- arr_flights: Broj dolaznih letova.
- arr_del15: Broj letova sa kašnjenjem od 15 minuta ili više.
- carrier_ct: Broj letova povezan s prevoznikom (kašnjenje zbog prevoznika).
- weather_ct: Broj letova povezan s vremenskim uslovima (kašnjenje zbog vremenskih uslova).
- nas_ct: Broj letova povezan s Nacionalnim vazduhoplovnim sistemom (NAS) (kašnjenje zbog NAS-a).
- security_ct: Broj letova povezan s bezbjednošću (kašnjenje zbog bezbjednosti).

- `late_aircraft_ct`: Broj letova povezan s kašnjenjem dolaska aviona (kašnjenje zbog kašnjenja dolaska aviona).
- `arr_cancelled`: Broj otkazanih letova.
- `arr_diverted`: Broj preusmjerenih letova.
- `arr_delay`: Ukupno kašnjenje dolaska.
- `carrier_delay`: Kašnjenje pripisano prevozniku.
- `weather_delay`: Kašnjenje pripisano vremenskim uslovima.
- `nas_delay`: Kašnjenje pripisano NAS-u.
- `security_delay`: Kašnjenje pripisano bezbjednosti.
- `late_aircraft_delay`: Kašnjenje pripisano kašnjenju dolaska aviona.

Adresa sa kog su preuzeti podaci se nalazi na sljedećem linku: [podaci](#).

• Algoritmi

Algoritmi koji će se koristiti za prediktivno modeliranje kašnjenja:

1. Linearna Regresija je statistički metod koji se koristi za modeliranje odnosa između zavisne promenljive (ciljne) i jedne ili više nezavisnih promenljivih (predictora). Cilj je pronaći linearnu vezu koja najbolje opisuje podatke.
2. Random Forest Regressor je ensemble model koji koristi više drveta odlučivanja (decision trees) kako bi postigao bolju prediktivnu tačnost i stabilnost u poređenju s pojedinačnim drvetom. Svako drvo u "šumi" donosi svoje predviđanje, a konačni rezultat se dobija agregiranjem predviđanja svih drveća. U kontekstu regresije, Random Forest Regressor koristi ovu tehniku kako bi predvideo numeričke vrijednosti.

• Tehnologije

Programski jezik koji je odlučeno da se koristi je Python. Python se često koristi za analizu podataka zbog biblioteka poput Pandas, NumPy, Scikit-learn, TensorFlow/Keras, Matplotlib i Seaborn.

Pandas kao biblioteka pruža mogućnost da se podaci učitaju i da se kreira data frame. Dok sklearn omogućava kreiranje modela koji će vršiti predikciju sa gore navedenim algoritmima. Matplotlib, Seaborn, Plotly i D3.js omogućavaju vizualizaciju podataka. Okruženje koje će se koristiti za izradu projekta je Visual Studio Code.

Flask će biti korišćen za backend i HTML/CSS/JavaScript za frontend. Kreiraće se korisnički interfejs za unos parametara i backend-a za generisanje predviđanja korišćenjem treniranih modela.

• Cilj

Cilj projekta jeste provesti detaljnu analizu podataka i omogućiti prediktivno modeliranje kašnjenja letova. Ovaj projekat ima za svrhu pružanje informacija korisnicima koje

će im pomoći u donošenju odluka vezanih za buduća putovanja, dok istovremeno doprinosi unapređenju operativnih performansi avio-kompanija.

Projekat će obuhvatati: analizu performansi prevoznika, identifikaciju trendova, analizu osnovnih uzroka kašnjenja, poređenje prevoznika, prediktivno modeliranje, kao i pridonošenje razumevanju industrije. Takođe, sistem će automatski generisati izvještaje i vizualizacije relevantnih informacija kako bi olakšao donošenje odluka i omogućio informisano upravljanje u avioindustriji.

Dok korisnici istražuju i analiziraju skup podataka putem intuitivnog korisničkog interfejsa, sticaće vrijedna saznanja koja će informisati procese donošenja odluka, unaprijediti operativne strategije i doprinijeti efikasnijem i pouzdanijem iskustvu putovanja avionom.