

Predviđanje otkazivanja hotelskih rezervacija

Projektni prijedlog

Kolegij: Strojno učenje

Profesor: dr.sc. Tomislav Šmuc

Asistenti: Matija Piškorec, Tomislav Lipić

Tim: Slučajni tim

Studenti: Matko Soče, Vinko Bokšić, Anđela Matijašević

Prirodoslovno matematički fakultet,

Sveučilište u Zagrebu,

Travanj 2020

1 Uvodni opis problema

Hotelska industrija temelji se na rezervacijama, a precizna procjena otkaza rezervacije je ključna za procjenu potražnje. Kako bi se umanjile posljedice otkazivanja rezervacija, hotelski menadžment provodi stroge politike otkaza i taktike "prekomjernog rezerviranja", što može imati negativne posljedice u smislu pada prihoda i loše reputacije hotela.

Kao što se može predvidjeti, stopa otkazivanja rezervacija u online industriji je prilično velika, a jednom kad se rezervacija otkáže gotovo ništa više se ne može učiniti, iz toga slijedi želja mnogih institucija za poduzimanjem mjera opreza, a predviđanje rezervacija koje mogu biti otkazane predstavlja dodatnu vrijednost hotelskoj industriji na temelju koje se mogu razraditi poslovne taktike s ciljem minimiziranja poslovnih gubitaka.

2. Cilj i hipoteze istraživanja

Korištenjem Kaggle skupa podataka želimo napraviti model koji će za svaku rezervaciju reći je li kandidat za otkazivanje ili ne.

Cilj je predvidjeti:

- broj otkazanih rezervacija u nekom razdoblju
- kolika je vjerojatnost da će određena rezervacija biti otkazana

Hipoteze:

- dio atributa određuje rezervaciju kao kandidata za otkazivanje
- na temelju danih podataka može se napraviti dovoljno dobar prediktivni model s preciznošću većom od 90%.

3 Pregled dosadašnjih istraživanja

U dosadašnjim istraživanjima razvijen je prototip sustava temeljen na strojnom učenju. Koriste se podaci hotelskih sustava za upravljanje imovinom i treniraju se klasifikacijski modeli kako bi se predvidjelo koje rezervacije će se otkazati i izračunala potražnja. Rezultati ukazuju na dobre performanse prototipa i daju važne pokazatelje za napredak u istraživanju i dokazuju da rezervacije s kojima hoteli kontaktiraju otkazuju manje od rezervacija koje nisu kontaktirane.

1. Istraživanje koje s više od 88% točnosti predviđa hoće li rezervacija biti uspješno realizirana ili otkazana. Hipoteza da je vrijeme rezervacije povezano s otkazivanjem rezervacije postaje vodeća prediktivna značajka modela. Druga važna značajka je kanal rezervacije, oni koji rezerviraju izravno ili putem putničkih agencija će s većom vjerojatnošću zadržati svoje rezervacije od onih koji rezerviraju putem internetskih stranica. Izlažući pokretače otkazivanje, modeli strojnog učenja mogu znatno pomoći hotelijerskoj industriji da bolje razumije obrasce ponašanja potencijalnih gostiju prema karakteristikama njihovih rezervacija.

2. U istraživanju su se koristila četiri skupa podataka za četiri različita hotela. Utvrđeno je da je tridesetak značajki dovoljno za napraviti dobar model predviđanja, a svi modeli navedeni u istraživanju dostigli su vrijednosti iznad 90%, čime je pokazano da su stabla odlučivanja uz pravilno identificiranje značajki dobar odabir algoritama za rješenje navedenog problema. U istraživanju je potvrđeno da jedan model predviđanja ne može biti primjenjiv na više hotela. Dodatno je navedeno kako razvoj poslovnog modela hotelskog menadžmenta treba biti popraćen i od strane modela predviđanja, jer svaka promjena i unaprjeđenje poslovnog procesa može utjecati na preciznost modela.

4 Materijali, metodologija i plan istraživanja

4.1 Na koji način će se probati riješiti problem

Ideja je riješiti problem metodama nadziranog učenja. Htjeli bismo iskoristiti neke od načina validacije modela i procjene preciznosti na test podacima izdvojenima iz navedenog skupa podataka.

4.2 Kako će se prikupiti podaci

Skup podataka preuzet s <https://www.kaggle.com/jessemostipak/hotel-booking-demand>

Opisana su dva skupa podataka, Resort hotel (H1) i City hotel (H2), oba hotela nalaze se u Portugalu, smještajnih kapaciteta više od 200 soba i klasificirani su kao hoteli s četiri zvjezdice.

Oba skupa imaju istu strukturu podataka, a dani podaci su za razdoblje od 01. srpnja 2015. do 31. kolovoza 2017. godine, obuhvaćaju realizirane i otkazane rezervacije

Podatke ćemo podijeliti na trening i test podatke.

4.3 Koje metode, algoritme, alate, tehnike će se koristiti?

Klasifikacija, kao oblik nadzirane tehnike strojnog učenja često se koristi za dobivanje predviđenih rezultata iz odabranog skupa podataka. Postoji mnogo algoritama za klasifikaciju u strojnom učenju: Randomforest, K Nearestneighbours, XGBoost, AdaBoost, GradientBoost, Neural Network, očekujemo da će naši podaci biti prikladniji korištenje više navedenih algoritama čiju učinkovitost ćemo uspoređivati, a model ćemo graditi koristeći unakrsnu validaciju.

Tehnike i alati koje planiramo koristiti: Jupyter (jezik Python) s paketima Pandas, Scikit-learn, Tensorflow

4.4. Kako će se ocijeniti uspješnost rezultata projekta?

Uspješnost rezultata projekta ocijeniti ćemo na temelju preciznosti dobivenih rezultata u odnosu na test podatke.

5 Očekivani rezultati predloženog projekta

Nadamo se da ćemo napraviti model pomoću kojeg će se na temelju odabranih atributa postići točnost od preko 90% na testnom skupu podataka.

6. Popis literature i reference

<https://web.math.pmf.unizg.hr/nastava/su/materijali/>

<https://www.kaggle.com/marcuswingen/eda-of-bookings-and-ml-to-predict-cancelations>

<file:///C:/Users/Korisnik/Downloads/Dialnet-PredictingHotelBookingCancellationsToDecreaseUncer-5921001.pdf>

<https://www.sciencedirect.com/science/article/pii/S2352340918315191>

<https://towardsdatascience.com/predicting-hotel-cancellations-with-machine-learning-fa669f93e794>

(1) <https://medium.com/@spettiett/hotel-booking-demand-insights-ba4e67eb287d>

(2) <https://pdfs.semanticscholar.org/0f5f/3a506360b9be0a7ab52d77974695f1c48a4d.pdf>