Fakultet elektrotehnike, računarstva i informacijskih tehnologija Osijek

Računarstvo usluga i analiza podataka

SEMINARSKI RAD

„Klasifikacija zauzeća prostorije na temelju nekoliko pokazatelja“

Kristijan Birtić

Luka Florschütz

Osijek, 2016.

Sadržaj

[1. Uvod 1](#_Toc461027559)

[2. Opis problema 2](#_Toc461027560)

[2.1. Korišteni podaci 2](#_Toc461027561)

[2.2. Korišteni postupci strojnog učenja 3](#_Toc461027562)

[3. Opis programskog rješenja 4](#_Toc461027563)

[3.1. Model strojnog učenja 4](#_Toc461027564)

[3.2. Način korištenja API-ja 6](#_Toc461027565)

[3.3. Klijentska aplikacija 8](#_Toc461027566)

[3.4. Dodatno 8](#_Toc461027567)

[4. Zaključak 9](#_Toc461027568)

[5. Poveznice i literatura 10](#_Toc461027569)

# Uvod

Zadatak ovog projekta je bio napraviti program koji će predvidjeti zauzeće prostorije na osnovu nekoliko parametara. Potrebno je preuzeti skup podataka, jedan dio iskoristiti za stvaranje klasifikacijskog modela, a drugi dio za treniranje i vrednovanje.

Nakon toga treba napraviti API koji će korisnicima ponuditi funkcionalnost našeg modela i izraditi klijentsku mobilnu/web/desktop aplikaciju koja će konzumirati API.

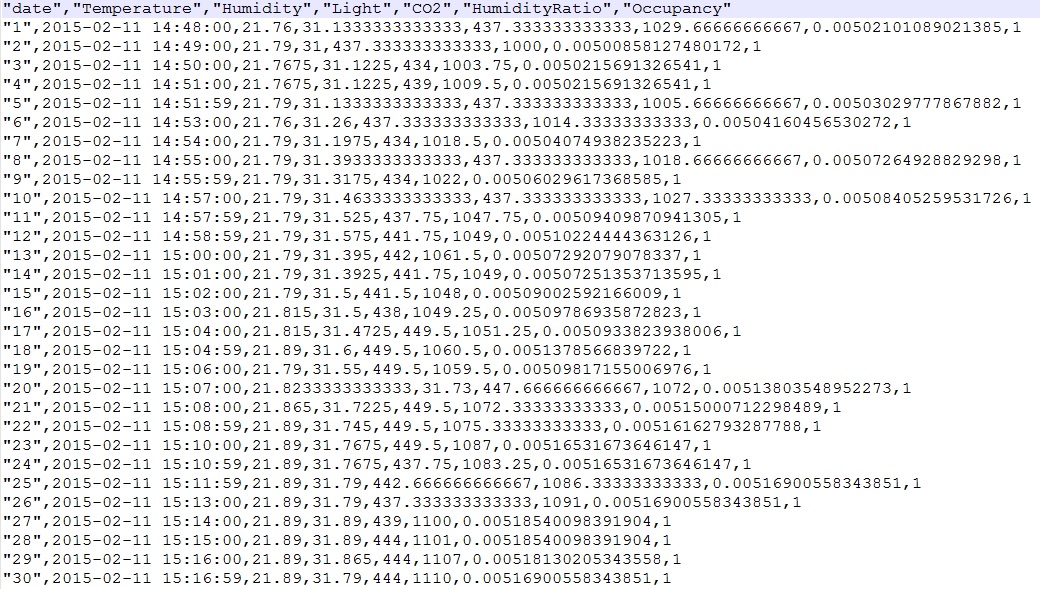
Cilj ovog projekta je omogućiti korisnicima uslugu predviđanja zauzeća prostorije za zadane parametre.

# Opis problema

Zauzeće prostorije se određuje logički sa 1 (prostorija je zauzeta), odnosno s 0 (prostorija je prazna). Parametri koje treba uzeti u obzir prilikom određivanja zauzeća prostorije su datum i vrijeme, temperatura, vlaga, svjetlost, razina CO2 i omjer vlage, te zauzeće koju će se morati predvidjeti. Računalu su dane tisuće podataka sa svim parametrima da bi se utreniralo, a onda mora na osnovu novih podataka odrediti zauzeće prostorije.

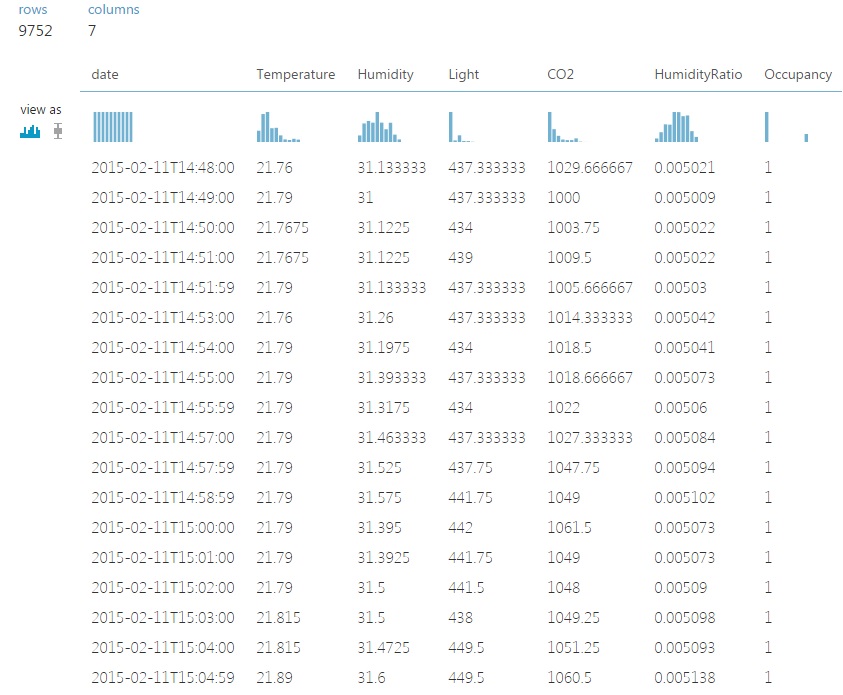
## Korišteni podaci

Korišteni podaci su preuzeti s UCI repozitorija [1]. Podaci su bili u .txt formatu, ali zapisani kao CSV. Samo je trebalo promjeniti format datoteke u .csv. Postojao je stupac viška kojeg je trebao obrisati prije uploadanja na Azure.



**Slika 1. Primjer neuređenog izvornog dataseta**

Na slici 1 dan je primjer neuređenog seta podataka. Postojalo je 7 naslovnih stupaca, ali 8 stupaca podataka. Višak je bio prvi stupac koji je označavao redni broj podataka. Kod uploadanja na Azure je bio problem što se prvi stupac interpretirao kao datum, drugi kao temperatura itd., pa su npr. podaci za svjetlost imali vrijednosti od podataka za vlagu itd. Zato je bilo potrebno ukloniti prvi stupac s rednim brojevima.



**Slika 2. Primjer dataseta**

Na slici 2. je dan primjer korištenog dataseta. Parametri su datum i vrijeme u danu, temperatura prostorije, vlaga u zraku, količina svjetlosti, količina CO2, omjer vlage i zauzeće. Sveukupno je 9752 redova podataka od kojih je 75% korišteno za treniranje modela, a ostalih 25% za evaluaciju.

## Korišteni postupci strojnog učenja

Strojno učenje je vrsta obrade podataka gdje se računalo može utrenirati da samo donosi neke rezultate ili odluke bez da je to eksplicitno isprogramirano. Postoji nadzirano, djelomično nadzirano, nenadzirano i učenje s povratnom vezom. Postoje razredi učenja, a to su klasifikacija, regresija, rangiranje, grupiranje i smanjenje dimenzionalnosti.

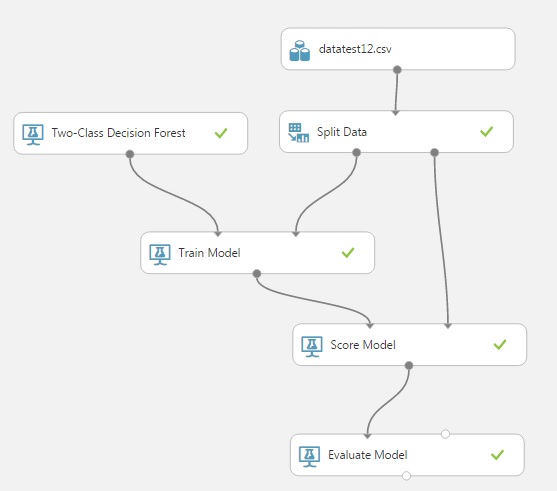
U projektu je korišeno nadzirano učenje. To znači da su računalu dani ulazi i izlazi, te ono na osnovu toga uči. Kao razred učenja korištena je klasifikacija. Ona radi tako što svakoj stavci dodjeljuje kategoriju.

# Opis programskog rješenja

Zadatak je bio napraviti program koji će na osnovu unesenih parametara prostorije kao izlaz vratiti predviđeno zauzeće. Program je napravljen u C#, dok je za strojno učenje korišten Azure Machine Learning Studio. Korisnik unosi parametre u program. Program se spaja na Azure, te šalje te podatke. Nakon što dobije odgovor u obliku stringa, korisniku ispisuje predviđeno zauzeće.

## Model strojnog učenja

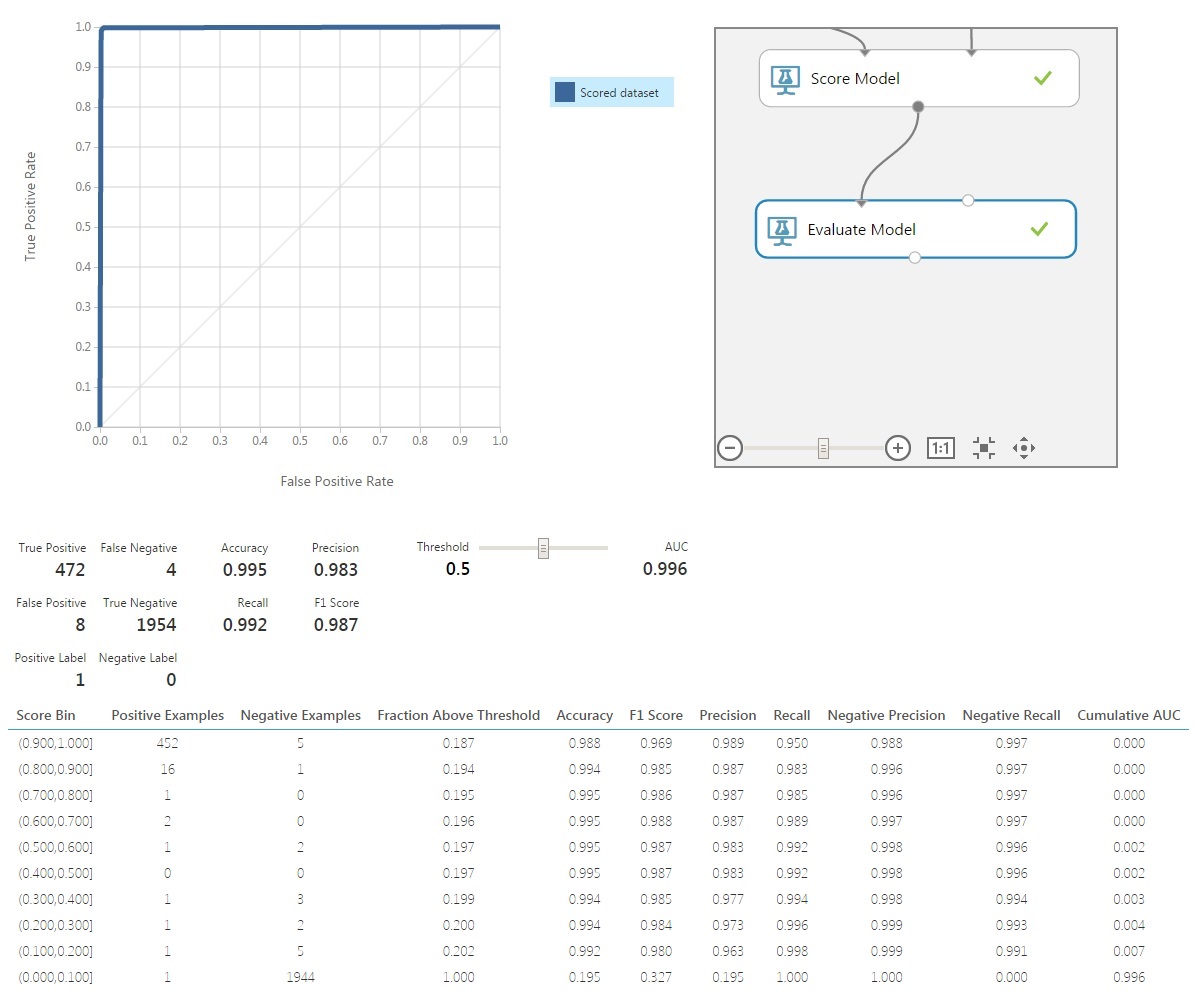
Model strojnog učenja vrlo je sličan onome što smo radili na laboratorijskim vježbama. Prvo smo upladali dataset, a nakon toga sastavili model. Odabrali smo 75% podataka za treniranje i 25% za evaluaciju.



**Slika 3. Azure ML eksperiment**

Za modul klasifikacije odabrali smo Two-Class Decision Forest. To je jedan od nekoliko tipova slučajne šume. Takve šume se sastoje od više stabala odlučivanja, te su učinkovite u računanju, treniranju i testiranju.

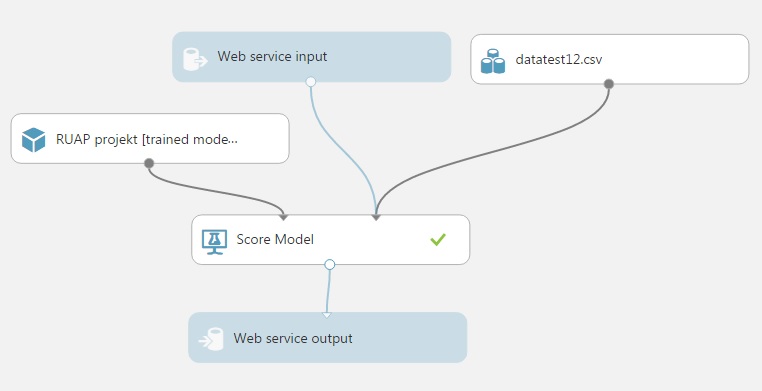
Two-Class Decision Forest modul služi za predviđanje parametra koji ima dvije vrijednosti. To je pogodno za ovaj projekt budući da se zauzeće prostorije iskazuje samo s dvije vrijednosti, tj. 1 za zauzeto i 0 za nije zauzeto.

**Slika 4. Evaluacija modela**

Ovaj model se pokazao jako dobrim. Ima vrlo visoke postotke točnosti, preciznosti, recall-a i F1 uspjeha. Za evaluaciju modela je korišteno oko 2400 podataka gdje je krivo predviđeno samo 12 rezultata.

## Način korištenja API-ja

Nakon što je model istreniran i pruža zadovoljavajuće rezultate, potrebno ga je staviti na Azure Web Service.



**Slika 5. Model za postavljanje na Web Service**

Time je napravljen API preko kojega se može putem Weba koristiti istrenirani model. Također je moguće preuzeti programski kod za klijentsku aplikaciju koji koristi taj API.

Uz taj programski kod dokumentirani su input i output. Input se sastoji od naziva stupaca i jednog ili više seta podataka. Kao output se vraćaju nazivi stupaca zajedno s tipovima podataka, unešenim vrijednostima i 2 nova podatka, scored labels i scored probabilities. Scored labels nam pokazuje vrijednost koju je model predvidio za zauzeće prostorije, a scored probabilites pokazuje vjerojatnost da je rezultat 1. Kako input i output izgleda se može vidjeti na slici 6.



**Slika 6. Primjeri inputa i outputa**

## Klijentska aplikacija

Kostur aplikacije je bio dan od strane Azurea u C#. Aplikaciju smo prilagodili tako da prima unos od korisnika, šalje zahtjev na Azure i dobija string kao odgovor.

Zbog jednostavnosti iz stringa smo izdvojili zadnja dva podatka, a to su scored label i probability. Na slikama 7 i 8 može se vidjeti kako program izgleda i radi. Korišteni su nasumični podaci iz danih datasetova.



**Slika 7. Klijentska aplikacija**



**Slika 8. Klijentska aplikacija**

# Zaključak

Prilikom izrade projektnog zadatka koristili smo C# i Microsoft Azure. Nakon umetanja podataka u Microsoft Azure i kreiranja ML modela , dobili smo kostur API-a kojeg smo oblikovali na svoj način tako da korisniku omogućimo konzolnu aplikaciju i lakše rukovanje. Korisnik unosi željene podatake, program ih predaje Azure-u, te kao povratnu informaciju dobiva predikciju u obliku stringa. Kako bi poboljšali predikciju, potrebno je više informacija.

# Poveznice i literatura

Programskom je rješenju moguće pristupiti preko:

|  |
| --- |
| https://github.com/BirticFlorschutz/Projekt |

1. K.P.Murphy, Machine Learning: A probabilistic perspective, MIT Press, Cambridge, Massachusetts, SAD, 2012
2. Labaratorijske vježbe