

Fakultet elektrotehnike, računarstva i informacijskih tehnologija Osijek

Računarstvo usluga i analiza podataka

SEMINARSKI RAD

„Predviđanje dijabetesa“

Marinela Bekić

Andrej Damašek

Osijek, 2026.

Sadržaj

1. UVOD.....	1
2. OPIS PROBLEMA	2
2.1. Korišteni podaci	2
2.2. Analiza podatkovnog skupa	3
2.3. Obrada podatkovnog skupa.....	6
2.4. Korišteni postupci strojnog učenja.....	8
3. Opis programskog rješenja	10
3.1. Model strojnog učenja	10
3.2. Način korištenja API-ja	13
3.3. Klijentska aplikacija	14
4. Zaključak.....	19
5. Poveznice i literatura.....	20

1. UVOD

Šećerna bolest ili dijabetes je dugotrajno stanje povišene razine glukoze u krvi, obilježeno poremećajima u metabolizmu ugljikohidrata, proteina i masti. Zbog njezina dugotrajnog trajanja mogu se razviti oštećenja velikih i malih krvnih žila, živaca te bazalnih membrana u različitim tkivima. Osobe oboljele od šećerne bolesti imaju povećan rizik od razvoja kardiovaskularnih bolesti, pri čemu dislipidemija, odnosno poremećaj masnoća u krvi, taj rizik dodatno pojačava u većoj mjeri nego u općoj populaciji [1].

Dijagnoza šećerne bolesti postavlja se na temelju određenih laboratorijskih pokazatelja. Dijagnostički kriteriji uključuju:

- HbA1c $\geq 6.5\%$
- Koncentracija glukoze u plazmi natašte $\geq 7,0$ mmol/L
- Vrijednost glukoze $\geq 11,1$ mmol/L tijekom testa oralnog opterećenja glukozom
- Slučajna razina glukoze u krvi $\geq 11,1$.

Razina glukoze u krvi može biti povišena u odnosu na normalne vrijednosti, ali još uvijek nedovoljna za postavljanje dijagnoze šećerne bolesti. U takvim se slučajevima govori o poremećenoj toleranciji glukoze, odnosno predijabetesu.

Cilj liječenja šećerne bolesti je produžiti život i poboljšati kvalitetu života oboljelih smanjenjem smrtnosti te sprječavanjem ili odgađanjem kroničnih komplikacija poput ateroskleroze, retinopatije, nefropatije i neuropatije. Temeljna načela liječenja uključuju uravnoteženu prehranu, tjelesnu aktivnost, edukaciju bolesnika, samokontrolu glukoze u krvi te primjenu odgovarajućih lijekova [1].

Analizom podataka o pacijentu moguće je, pomoću određenih indikatora, procijeniti rizik od razvoja šećerne bolesti. U praksi takva procjena nije uvijek izvediva za sve pacijente zbog ograničenog broja zdravstvenog osoblja i nedostatne pažnje posvećene ranoj identifikaciji rizika. Cilj ovog seminarskog rada je nadomjestiti te nedostatke razvojem jednostavne aplikacije koja na temelju dostupnih podataka procjenjuje rizik pojave šećerne bolesti. U radu se opisuje podatkovni skup korišten za izradu modela strojnog učenja, postupak njegove pripreme te analiza odnosa između pojedinih varijabli uz pomoć grafičkih prikaza. Nadalje, prikazan je proces odabira najprikladnijeg algoritma, vrednovanje njegovih performansi te implementacija i primjena modela u okviru okruženja *Microsoft Azure* [2]. Također je razvijena web aplikacija namijenjena jednostavnom unosu podataka, prikazu rezultata predviđanja i osnovnoj statistici vezanoj uz korišteni podatkovni skup.

2. OPIS PROBLEMA

Šećerna bolest (dijabetes) predstavlja ozbiljan zdravstveni problem koji pogađa sve veći broj ljudi širom svijeta. Rana identifikacija osoba s visokim rizikom od razvoja dijabetesa ključna je za pravovremenu prevenciju i smanjenje komplikacija. Međutim, u praksi je teško sustavno pratiti sve pacijente zbog ograničenog broja zdravstvenih djelatnika i velikog volumena podataka.

Cilj ovog projekta je razviti sustav koji koristi dostupne podatke o pacijentima; poput dobnihi podataka, tjelesne mase, vrijednosti glukoze, krvnog tlaka i drugih relevantnih parametara kako bi predvidio rizik od pojave dijabetesa. Takav sustav omogućuje ranije prepoznavanje visokorizičnih osoba, pomaže liječnicima u donošenju odluka i može služiti kao alat za prevenciju bolesti. Problem se može formalizirati kao klasifikacijski zadatak strojnog učenja, gdje model uči na povijesnim podacima i predviđa hoće li osoba razviti dijabetes ili ne.

2.1. Korišteni podaci

Za izradu modela strojnog učenja potreban je podatkovni skup koji omogućuje procjenu parametara i predviđanje stanja novih pacijenata. Odabrani skup podataka dostupan je na platformi Kaggle [3].

Skup se sastoji od 100.000 uzoraka, a sadrži slijedeće podatke o pacijentu:

- *gender*: spol pacijenta (muškarac ili žena)
- *age*: starost pacijenta
- *hypertension*: binarna vrijednost koja simbolizira postojanje hipertenzije kod pacijenta (1 ako ima, 0 ako nema)
- *heart_disease*: binarna vrijednost koja simbolizira postojanje bolesti srca kod pacijenta (1 ako ima, 0 ako nema)
- *smoking_history*: pokazuje je li pacijent ovisnik o cigaretama (nema informacija, trenutni pušač, ikad, prestao pušiti, nikad, nije trenutni pušač)
- *bmi*: indeks tjelesne mase
- *HbA1c_level*: vrijednost glikiranog hemoglobina, pokazuje prosječnu razinu glukoze u krvi pacijenta
- *blood_glucose_level*: trenutna razina glukoze u krvi pacijenta
- *diabetes*: binarna vrijednost koja simbolizira je li pacijent ima dijabetes ili ne (1 ako ima, 0 ako nema).

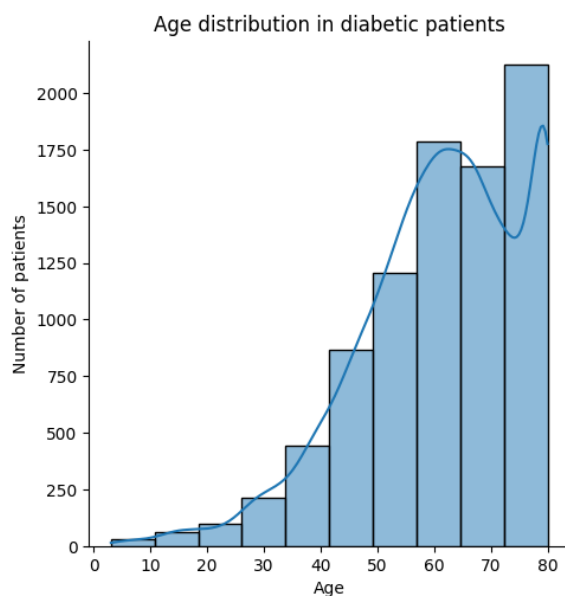
Na slici 1 prikazano je prvih 5 primjera podataka o pacijentima u tabličnom formatu. Podatkovni je skup preuzet u .CSV formatu te je obrađen u okviru *Google Colaboratory* unutar *Jupyter Notebook* datoteke [4] koja je pogodna za analizu skupova podataka te strojno učenje. Datoteka u kojoj se nalazi analiza skupa, kao i korišteni skupovi podatka se nalazi u mapi *Analiza* koja je dostupna u *GitHub* repozitoriju.

	gender	age	hypertension	heart_disease	smoking_history	bmi	HbA1c_level	blood_glucose_level	diabetes
0	Female	80.0	0	1	never	25.19	6.6	140	0
1	Female	54.0	0	0	No Info	27.32	6.6	80	0
2	Male	28.0	0	0	never	27.32	5.7	158	0
3	Female	36.0	0	0	current	23.45	5.0	155	0
4	Male	76.0	1	1	current	20.14	4.8	155	0

Slika 1 Primjeri podataka o pacijentima

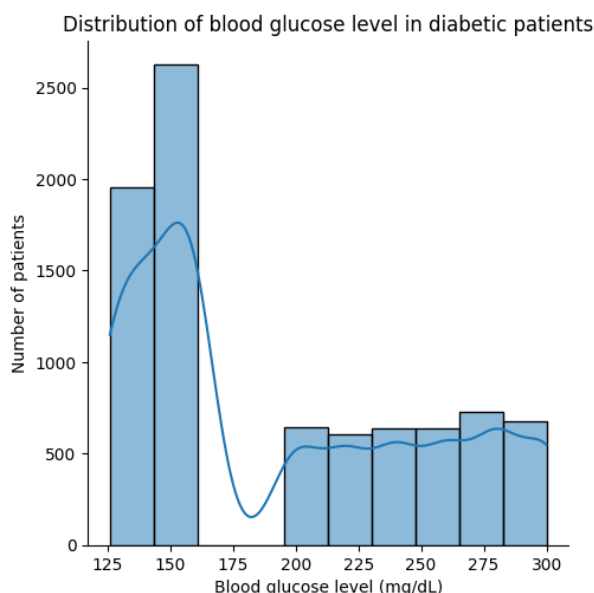
2.2. Analiza podatkovnog skupa

Pomoću analize moguće je utvrditi i vizualizirati odnose između značajki, njihovu distribuciju i povezanost. Unutar *Jupiter Notebook* datoteke kreirani su dijagrami kako bi se prikazala distribucija značajki te povezanost sa izlaznom varijablom.



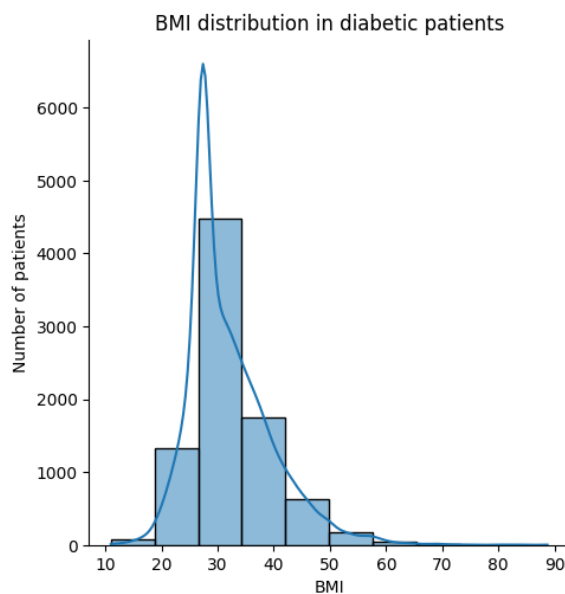
Slika 2 Distribucija pacijenata koji imaju dijabetes prema godinama starosti

Na slici 2 je prikazana distribucija pacijenata koji imaju dijabetes prema godinama starosti, vidljivo je da prevladavaju pacijenti nakon 55 godine života, dok mali broj pacijenata je u dobi između 0 i 35.



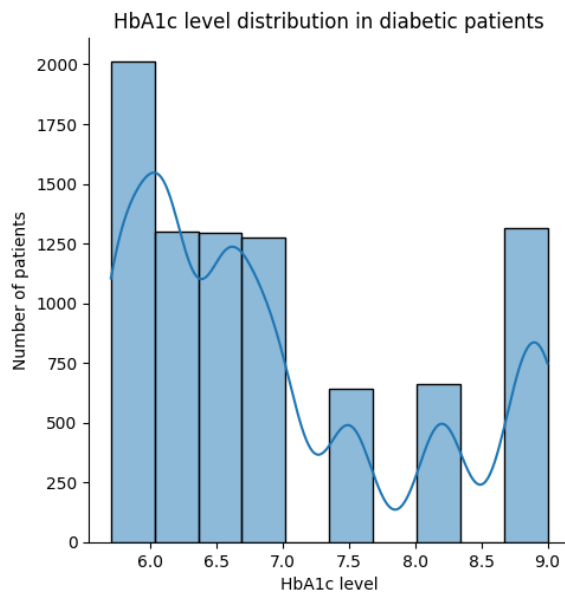
Slika 3 Distribucija pacijenata koji imaju dijabetes prema razini glukoze u krvi

Na slici 3 je prikazana distribucija pacijenata koji imaju dijabetes prema razini glukoze u krvi, vidljivo je da najveći broj pacijenata ima razinu glukoze oko 150 mg/dL, dok gotovo niti jedan pacijent s dijabetesom nema razinu glukoze između 150 i 200. Normalna razina glukoze u krvi iznosi između 70 i 99 mg/dL [5].



Slika 4 Distribucija pacijenata koji imaju dijabetes prema indeksu tjelesne mase

Na slici 4 je prikazana distribucija pacijenata koji imaju dijabetes prema indeksu tjelesne mase, vidljivo je da prevladavaju pacijenti sa indeksom mase između 25 i 35 što zapravo označava pretilost prvog stupnja [6].

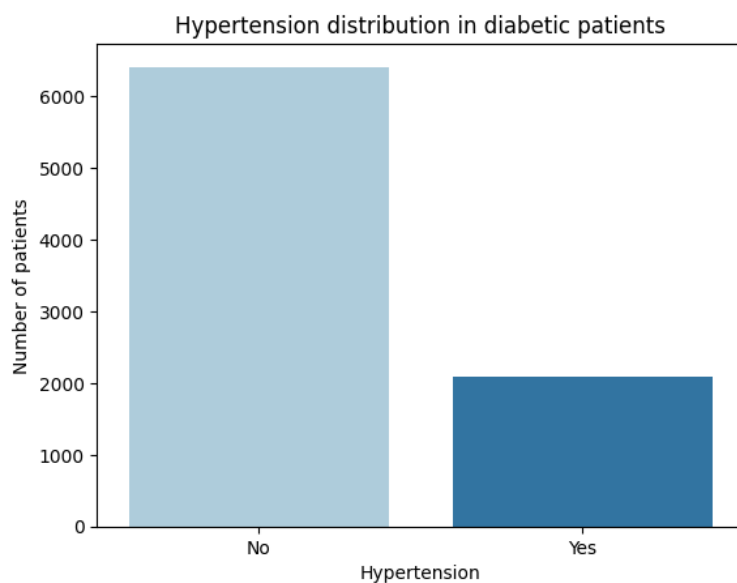


Slika 5 Distribucija pacijenata koji imaju dijabetes prema razini glikiranog hemoglobina

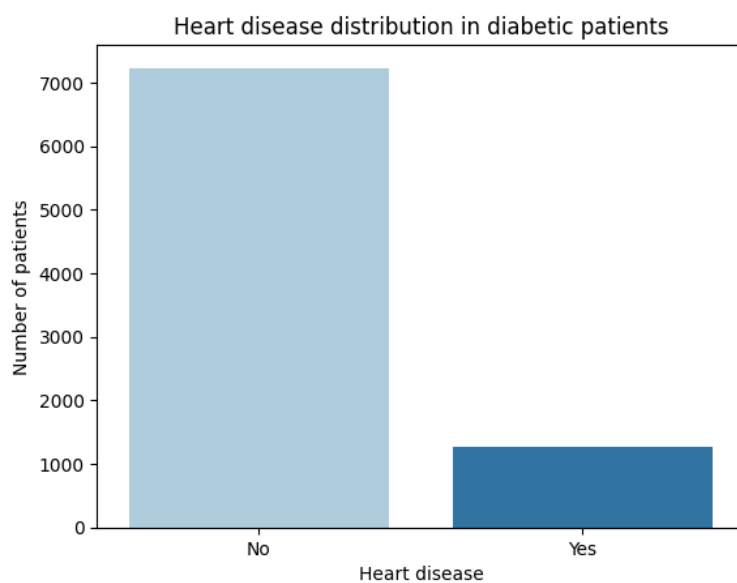
Na slici 5 je prikazana distribucija pacijenata koji imaju dijabetes prema razini glikiranog hemoglobina. Normalna razina HbA1c je ispod 5.7%, a prema distribuciji je vidljivo da pacijenti imaju razinu HbA1c iznad normalne razine [7].

Na slikama 6, 7 i 8 prikazan je odnos pacijenata s dijabetesom s obzirom na hipertenziju, srčanim bolestima te povijesti pušenja. Vidljivo da pacijenti koji imaju dijabetes u većini nemaju hipertenziju

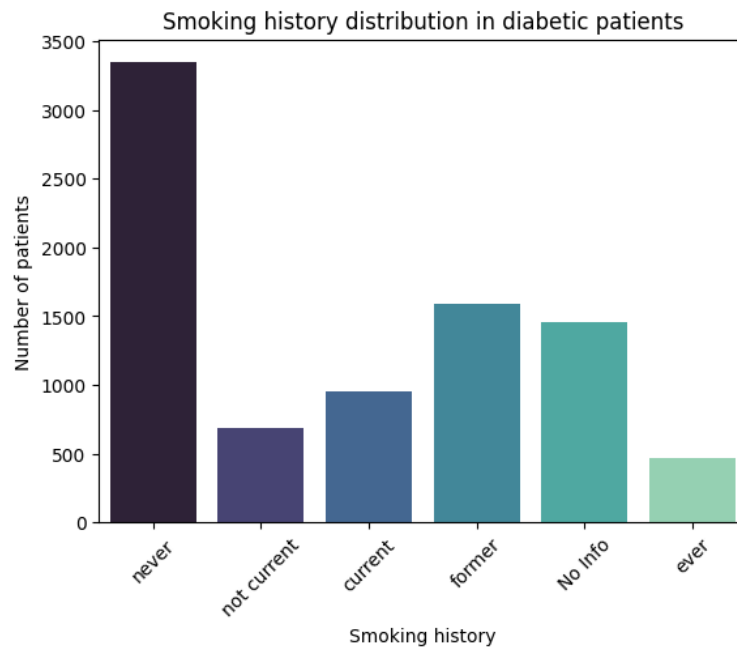
nit i srčane bolesti te se ovisnost o pušenju ne može povezati sa dijabetesom jer prevladavaju osobe koje nikada nisu pušile, a imaju dijabetes. Dok je na slici 9 prikazan odnos muških i ženskih pacijenata koji imaju dijabetes. Vidljivo je da su pacijenti s obzirom na spol podjednaki.



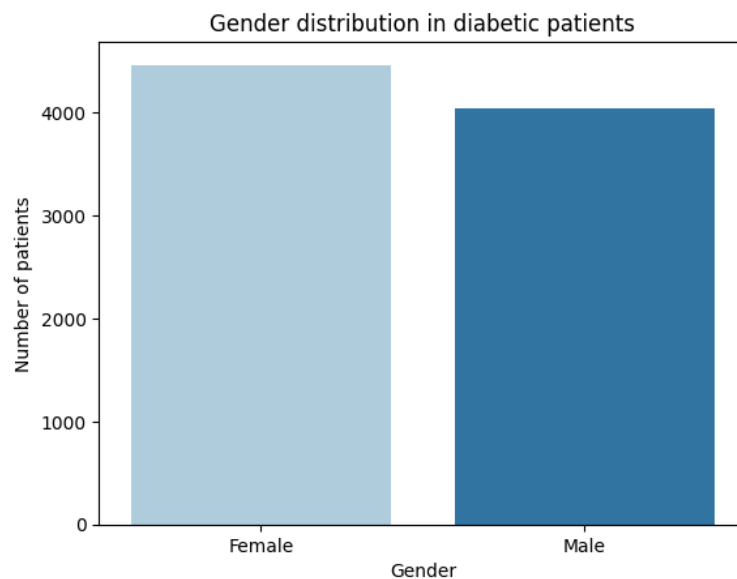
Slika 6 Odnos pacijenata koji imaju dijabetes sa hipertenzijom



Slika 7 Odnos pacijenata koji imaju dijabetes sa srčanom bolesti



Slika 8 Odnos pacijenata koji imaju dijabetes sa srčanom bolesti



Slika 9 Odnos ženskih i muških pacijenata koji imaju dijabetes

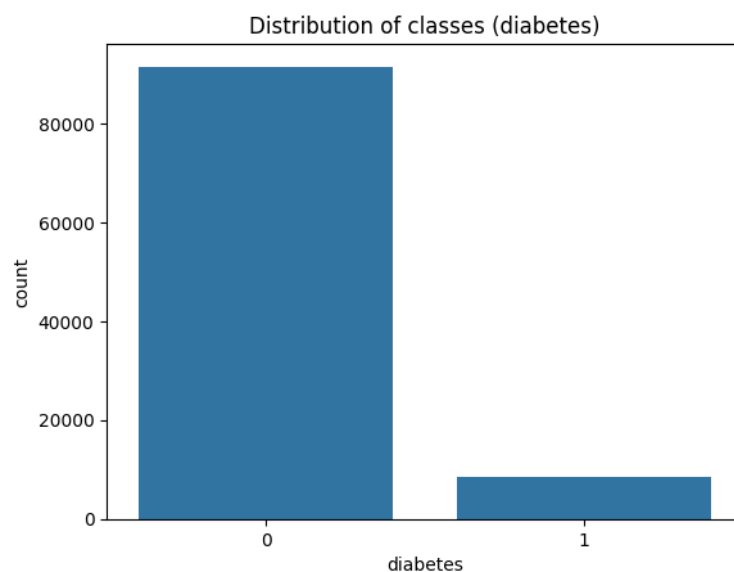
2.3. Obrada podatkovnog skupa

Obrada podataka je izuzetno važan proces kojim se podaci pripremaju za učenje modela. Potrebno je izvršiti transformacije nad određenim značajkama podatkovnog skupa kako bi se format podataka prilagodio algoritmima strojnog učenja. Tako je kategorička značajka *smoking_history* korištenjem *One-Hot Encoding* tehnike pretvorena u binarne vrijednosti, odnosno za svaku pojedinu vrijednost značajke stvorio se novi stupac koji sadrži vrijednost 0 ili 1. Također je potrebno vrijednosti značajke *gender* mapirati u 0 ili 1 kako bi se izbjegli mogući problemi pri obradi tekstualnih vrijednosti, a značajku *age* bi bilo dobro pretvoriti u *integer*, odnosno cijeli broj. Na slici 10 prikazane su opisane izmjene u skupu podataka.

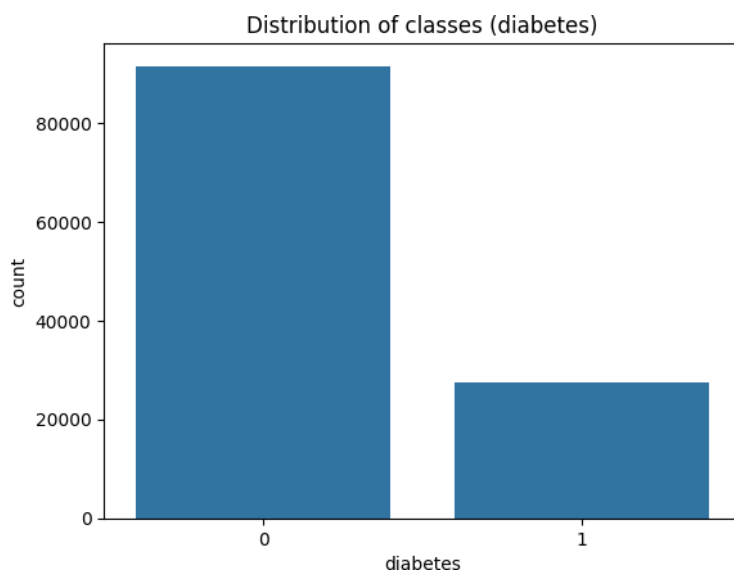
	gender	age	hypertension	heart_disease	bmi	HbA1c_level	blood_glucose_level	diabetes	No Info	current	ever	former	never	not current
0	0	80	0	1	25.19	6.6	140	0	0.0	0.0	0.0	0.0	1.0	0.0
1	0	54	0	0	27.32	6.6	80	0	1.0	0.0	0.0	0.0	0.0	0.0
2	1	28	0	0	27.32	5.7	158	0	0.0	0.0	0.0	0.0	1.0	0.0
3	0	36	0	0	23.45	5.0	155	0	0.0	1.0	0.0	0.0	0.0	0.0
4	1	76	1	1	20.14	4.8	155	0	0.0	1.0	0.0	0.0	0.0	0.0

Slika 10 Skup podataka nakon provedbe *One-hot Encodinga* i mapiranja

Na slici 11 prikazan je odnos između klasa pacijenata koji imaju dijabetes i oni koji nemaju. Vidljivo je da prevladavaju pacijenti bez dijabetesa, omjer pacijenata bez dijabetesa i sa dijabetesom iznosi 91482 : 8500 stoga je potrebno napraviti preuzorkovanje (engl. *Oversampling*) na način da se udio primjeraka manjinske klase u ukupnom skupu dovede do 30% kako ne bi došlo do favoriziranja većinske klase u modelu jer takav omjer daje kvalitetnije rezultate. Na slici 11 i 12 je prikazan odnos pacijenata sa dijabetesom i bez dijabetesa prije i nakon preuzorkovanja skupa podataka.



Slika 11 Odnos između pacijenata s dijabetesom i bez dijabe

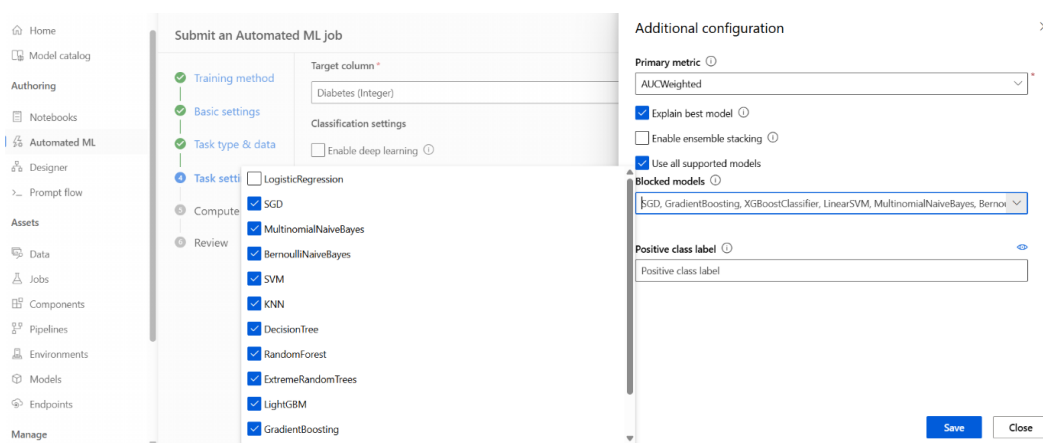


Slika 12 Odnos između pacijenata s dijabetesom i bez dijabetesa nakon preuzorkovanja skupa podataka

Nakon provedenog postupka obrade i preuzorkovanja, podatkovni skup pohranjen je u novu .CSV datoteku koja se koristi za učenje modela unutar okruženja *Microsoft Azure*, a nalazi se u istoj mapi kao i izvorna datoteka.

2.4. Korišteni postupci strojnog učenja

Prije izrade modela u okruženju *Microsoft Azure* potrebno odabrati odgovarajući algoritam strojnog učenja te optimizirati njegove hiperparametre. Fokus je stavljen na algoritam logističke regresije, jer je to robusna i često korištena metoda za binarne klasifikacijske probleme, poput predviđanja prisutnosti ili odsutnosti šećerne bolesti. Dodatna prednost logističke regresije je njezina interpretabilnost, koja omogućuje lakše razumijevanje utjecaja pojedinih ulaznih varijabli na konačnu odluku modela. Optimizacija hiperparametara provedena je korištenjem *AutoML* pristupa, kojim su određene najprikladnije vrijednosti za promatrani podatkovni skup. Prilikom vrednovanja modela korištena je K unakrsna validacija s vrijednošću $K = 5$ budući da ova metoda omogućuje pouzdaniju procjenu performansi modela i smanjuje rizik od prenaučenosti na trening-podatke. Podatkovni skup dodatno je podijeljen metodom *Train-Test*, pri čemu je 20 % podataka izdvojeno za testiranje modela. Na slikama 12 i 13 prikazane su postavke odabira algoritma, odnosno hiperparametara.



Slika 13 Odabir algoritma Logistička regresija

Validation type ⓘ
k-fold cross validation ▼

Number of cross validations * ⓘ
5 *

Test data ⓘ
Train-test split ▼

Percentage test of data * ⓘ
20 *
Automated ML recommends that between 10 and 30 percent of data is held out for test

Slika 14 Odabir provjere i podjele podataka

Nakon pokretanja *AutoML* postupka pronađen je model s najbolje ostvarenim rezultatima. Na temelju tog modela određeni su optimalni hiperparametri za daljnju izgradnju modela. Na slici 15 je prikaz nekoliko najboljih modela, a na slici 16 su prikazane postavke hiperparametara najboljeg modela, koje su zatim korištene pri izgradnji konačnog modela primjenom *Designer* pristupa u okruženju *Microsoft Azure*.

Algorithm name	Responsible AI	AUC weighted	Sampling	Created on	Duration	Hyperparameter
VotingEnsemble		0.97448	100.00 %	Feb 2, 2026 6:58 PM	1m 4s	algorithm : [LogisticRegression], ...
StandardScalerWrapper, LogisticRegression		0.97448	100.00 %	Feb 2, 2026 6:40 PM	46s	C : 3.727593720314938 class, ...
StandardScalerWrapper, LogisticRegression		0.97448	100.00 %	Feb 2, 2026 6:40 PM	33s	C : 0.040949150623804234 class, ...
StandardScalerWrapper, LogisticRegression		0.97448	100.00 %	Feb 2, 2026 6:40 PM	35s	C : 4714.8663634573895 class, ...
MaxAbsScaler, LogisticRegression		0.97447	100.00 %	Feb 2, 2026 6:40 PM	36s	C : 1.7575106248547894 class, ...
MaxAbsScaler, LogisticRegression		0.97447	100.00 %	Feb 2, 2026 6:40 PM	35s	C : 4714.8663634573895 class, ...

Slika 15 Prikaz najboljih modela

```

Training algorithm:
1  {
2      "class_name": "LogisticRegression",
3      "module": "sklearn.linear_model",
4      "param_args": [],
5      "param_kwargs": {
6          "C": 3.727593720314938,
7          "class_weight": "balanced",
8          "multi_class": "ovr",
9          "penalty": "l1",
10         "solver": "saga"
11     },
12     "prepared_kwargs": {},
13     "spec_class": "sklearn"
14 }

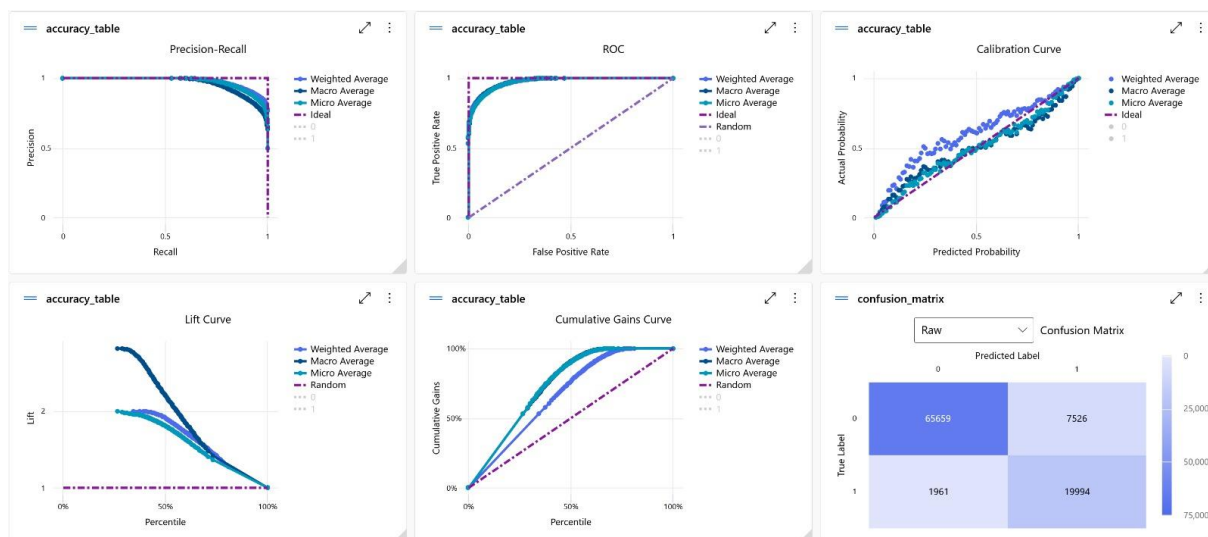
```

Slika 16 Prikaz hiperparametara odabranog modela

Na slici 17 su prikazani dobiveni rezultati i metrike odabranog modela, a na slici 18 grafovi i matrica zabune. Vidljivo je da su rezultati zadovoljavajući, stoga smo koristili hiperparametre toga modela (slika 16).

accuracy 0.9002838	AUC_macro 0.9744770	AUC_micro 0.9737786	AUC_weighted 0.9744770	average_precision_sco... 0.9628147	average_precision_sco... 0.9751070	average_precision_sco... 0.9785983	balanced_accuracy 0.9039204	f1_score_macro 0.8704389
f1_score_micro 0.9002838	f1_score_weighted 0.9039202	log_loss 0.2023301	matthews_correlation 0.7506748	norm_macro_recall 0.8078409	precision_score_macro 0.8487777	precision_score_micro 0.9002838	precision_score_weigh... 0.9145863	recall_score_macro 0.9039204

Slika 17 Metrike odabranog modela



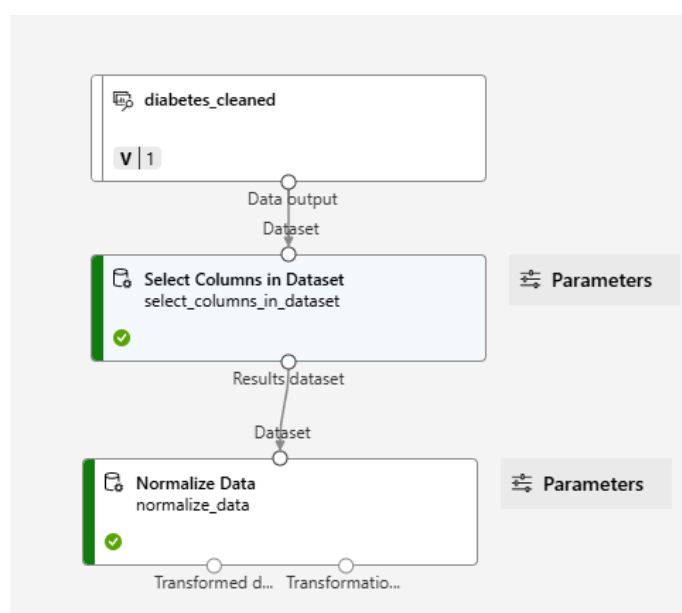
Slika 18 Grafovi i matrica zabune odabranog modela

3. Opis programskog rješenja

Nakon uspješnog pronalaska algoritma s najboljim hiperparametrima, u *Azure Machine Learning Studio* okruženju stvaramo potreban model koji se trenira na podacima te nakon treniranja postavljamo krajnju točku.

3.1. Model strojnog učenja

Model je izrađen korištenjem pristupa *Designer* u okruženju *Microsoft Azure*. Kao prvi korak uvezen je prethodno pripremljen podatkovni skup, koji je obrađen u *Jupyter Notebooku*. Nakon uvoza podataka korišten je modul *Select Columns in Dataset* kako bi imali kontrolu koje stupce ćemo koristiti u daljnjoj analizi, a zatim je provedena normalizacija podataka. Cjelokupni tijek obrade, odnosno cjevovod prikazan je na slici 19, dok je na slici 20 prikazan dobiveni podatkovni skup nakon pokretanja i izvršavanja cjevovoda kako bi se uvjerili da je uvoz i normalizacija podataka bila uspješna.

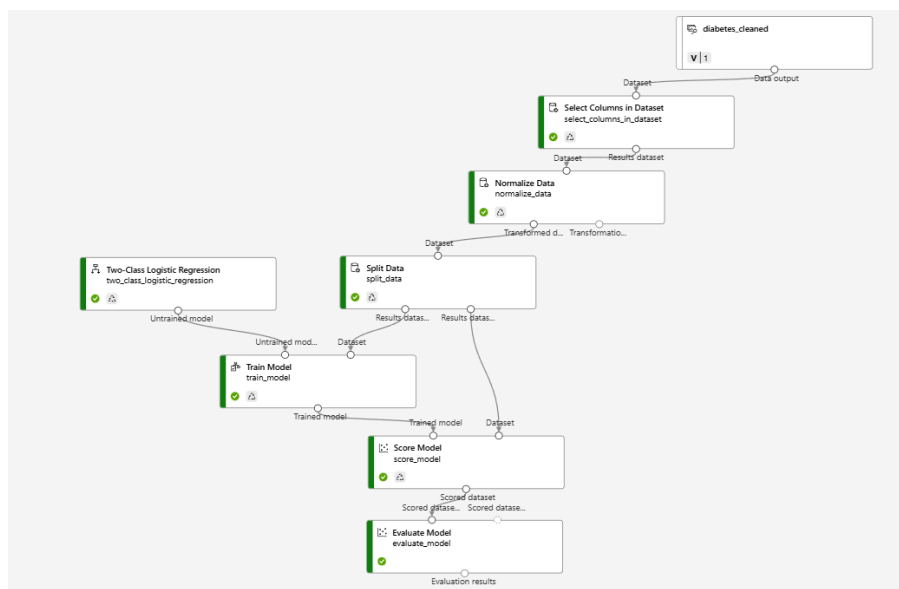


Slika 19 Prikaz cjevovoda za uvoz i normalizaciju podataka

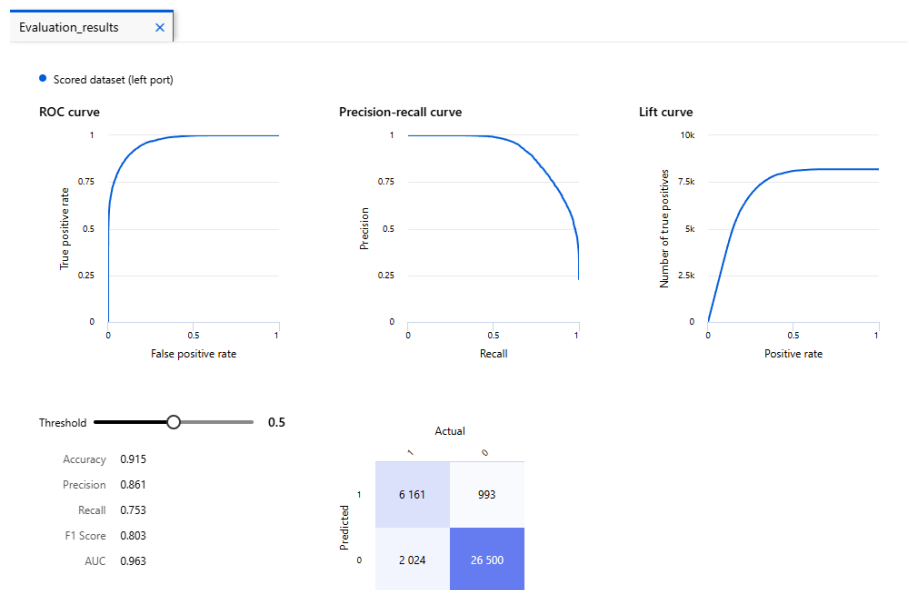
gender	age	hypertension	heart_disease	bmi	HbA1c_level	blood_glucose_level	No Info	current	ever	former	never	not current	Diabetes
0	1	0	1	0.177171	0.563636	0.272727	0	0	0	0	1	0	0
0	0.674675	0	0	0.202031	0.563636	0	1	0	0	0	0	0	0
1	0.349349	0	0	0.202031	0.4	0.354545	0	0	0	0	1	0	0
0	0.449449	0	0	0.156863	0.272727	0.340909	0	1	0	0	0	0	0
1	0.94995	1	1	0.118231	0.236364	0.340909	0	1	0	0	0	0	0
0	0.249249	0	0	0.202031	0.563636	0.022727	0	0	0	0	1	0	0
0	0.54955	0	0	0.108543	0.545455	0.545455	0	0	0	0	1	0	1
0	0.987487	0	0	0.161648	0.4	0.022727	1	0	0	0	0	0	0
1	0.524525	0	0	0.275794	0.236364	0.295455	0	0	0	0	1	0	0
0	0.399399	0	0	0.202031	0.272727	0.090909	0	0	0	0	1	0	0
0	0.662162	0	0	0.202031	0.472727	0.022727	0	0	0	0	1	0	0
0	0.674675	0	0	0.521592	0.454545	0.090909	0	0	0	1	0	0	0
0	0.974975	0	0	0.303922	0.272727	0.227273	0	0	0	1	0	0	0
0	0.837337	0	0	0.183007	0.418182	0.545455	0	0	0	0	1	0	0
0	0.94995	0	0	0.202031	0.272727	0.363636	1	0	0	0	0	0	0
1	0.974975	0	0	0.202031	0.563636	0.209091	1	0	0	0	0	0	0
1	0.186687	0	0	0.237512	0.472727	0.545455	0	0	0	0	1	0	0
0	0.524525	0	0	0.168884	0.4	0.354545	0	0	0	0	1	0	0
0	0.524525	0	0	0.202031	0.4	0	1	0	0	0	0	0	0
1	0.461962	0	0	0.183357	0	0.359091	0	0	1	0	0	0	0
1	0.499499	0	0	0.307773	0.454545	0.045455	0	1	0	0	0	0	0
1	0.061562	0	0	0.102591	0.490909	0.022727	1	0	0	0	0	0	0
0	0.862362	0	0	0.131069	0.236364	0.022727	0	0	0	0	1	0	0
0	0.8999	0	1	0.209267	0.545455	0.227273	0	0	0	1	0	0	0
0	0.049049	0	0	0.046452	0.090909	0.272727	1	0	0	0	0	0	0
1	0.374374	0	0	0.277194	0.472727	0.209091	0	0	0	0	1	0	0
1	0.837337	0	1	0.202031	0.545455	0.545455	0	0	0	0	0	1	1

Slika 20 Prikaz podataka nakon normalizacije

Nakon uspješnog uvoza i normalizacije podataka napravljen je model za treniranje najboljeg modela, a njegov cjevovod je prikazan na slici 21. Pomoću komponente *Split data* definirali smo da se 80% podataka koristi za treniranje, dok se ostalih 20% testira. Na slici 22 prikazani su grafovi, metrike i matrica zabune istreniranog modela koji su dobiveni pomoću komponente *Evaluate Model*.

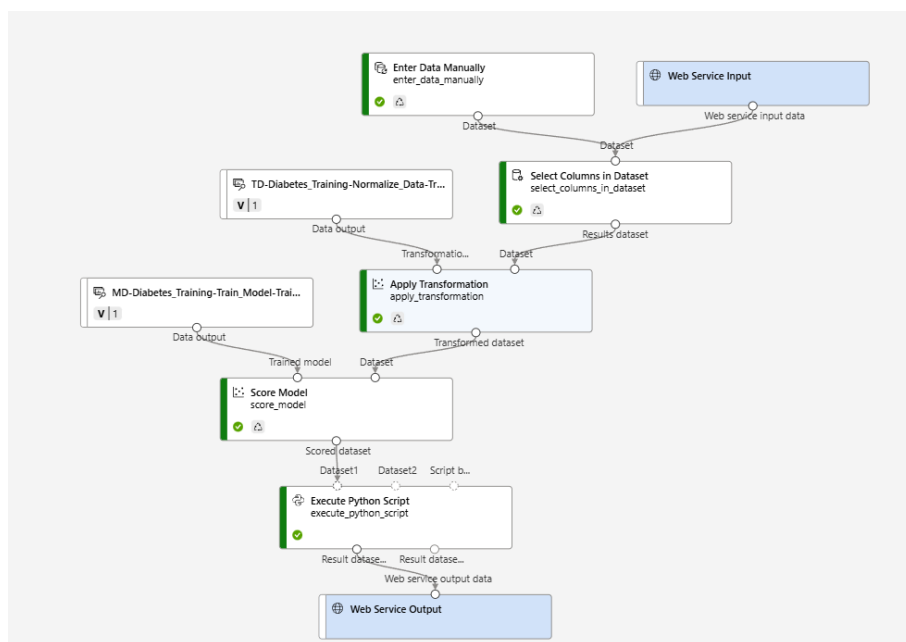


Slika 21 Cjevovod za treniranje modela



Slika 22 Rezultati istreniranoga modela

Na slici 23 se nalazi cjevovod koji koristi cjevovod kao komponentu sa slike 21 za unos i slanje podataka pomoću *web servisa*. Komponenta *Web Service Input* služi kako bi primili podatke sa *web aplikacije*, a *Web Service Output* za prikaz rezultata na klijentskoj aplikaciji. Pomoću komponente *Execute Python Script* dobivene rezultate iz komponente *Score Model* moguće je koristiti za daljnju upotrebu ili prikaz. Kod iz *Python* skripte je prikazan na slici 24.



Slika 23 Cjevovod za primanje i slanje podataka putem web servisa

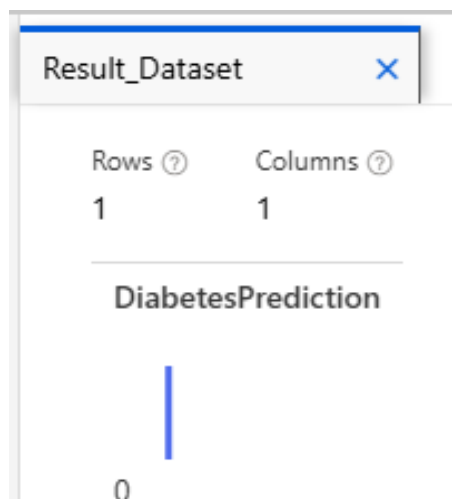
Execute Python Script

Python script ⓘ *

```
1 import pandas as pd
2
3 def azureml_main(dataframe1 = None, dataframe2 = None):
4     scored_results = dataframe1[['Scored Labels']]
5
6     scored_results.rename(columns={'Scored Labels':'DiabetesPrediction'},
7                             inplace=True)
8
9     return scored_results
```

Slika 24 Python kod za prikaz rezultata

Kako bi se provjerila ispravnost modela, u komponenti *Enter Data Manually* uneseni su podaci pacijenta bez dijabetesa: muška osoba u dobi od 37 godina, bez hipertenzije i bolesti srca, s indeksom tjelesne mase 25,72, razinom HbA1c od 3,5 te razinom glukoze u krvi od 159 mg/dL, uz povijest povremenog pušenja. Podaci su uneseni redoslijedom i odvojeni zarezima, bez uključivanja stupca Diabetes, u sljedećem obliku: 1, 37.0, 0, 0, 25.72, 3.5, 159, 0, 0, 1, 0, 0, 0. Kao rezultat dobivena je predikcija o prisutnosti ili odsutnosti dijabetesa kod navedenog pacijenta. Dobiveni i uspješni rezultati su prikazani na slici 25.



Rows ⓘ	Columns ⓘ
1	1

DiabetesPrediction
0

Slika 25 Rezultati modela

Nakon uspješne provjere ispravnosti cjevovoda stvara se kranja točka kako bi opisani cjevovodi bio dostupan za korištenje u *web* aplikaciji.

3.2. Način korištenja API-ja

API je baziran na REST arhitekturi, shodno tome koristi HTTP protokol za obradu zahtjeva. API je realiziran kao *web servis* pomoću *Azure Machine Learning Studio-a*.

U tablici 1. prikazani su podaci koji se šalju ili primaju kada se poziva krajnja točka realizirana pomoću *Azure Machine Learning Studio-a*. Krajnja točka se poziva pomoću POST zahtjeva na njezinu adresu. Tip sadržaja šalje se u zaglavlju. U tijelu zahtjeva navodi se JSON objekt koji sadrži sve potrebne stupce za obavljanje predviđanja. Kao odgovor dobiva se numerički rezultat o tome da li pacijent potencijalno ima dijabetes ili ga nema. Za predane podatke iz tablice 1. predviđanje je da osoba ima dijabetes.

Adresa krajnje točke	http://761e0046-f8d3-4485-af0d-7ff6f8b1f918.polandcentral.azurecontainer.io/score
Metoda	POST
Zaglavlje	'Content-Type':'application/json'
Tijelo zahtjeva	<pre>{ "Inputs": { "data": [{ "gender": 1, "age": 45, "hypertension": 0, "heart_disease": 0, "bmi": 27.5, "HbA1c_level": 5.8, "blood_glucose_level": 110, "No Info": 0, "current": 1, "ever": 0, "former": 0, "never": 0, "not current": 0 }] }, "GlobalParameters": { "method": "predict" } }</pre>
Tijelo odziva	<pre>{ "Results": [1] }</pre>

Tablica 1. Podaci vezani za krajnju točku

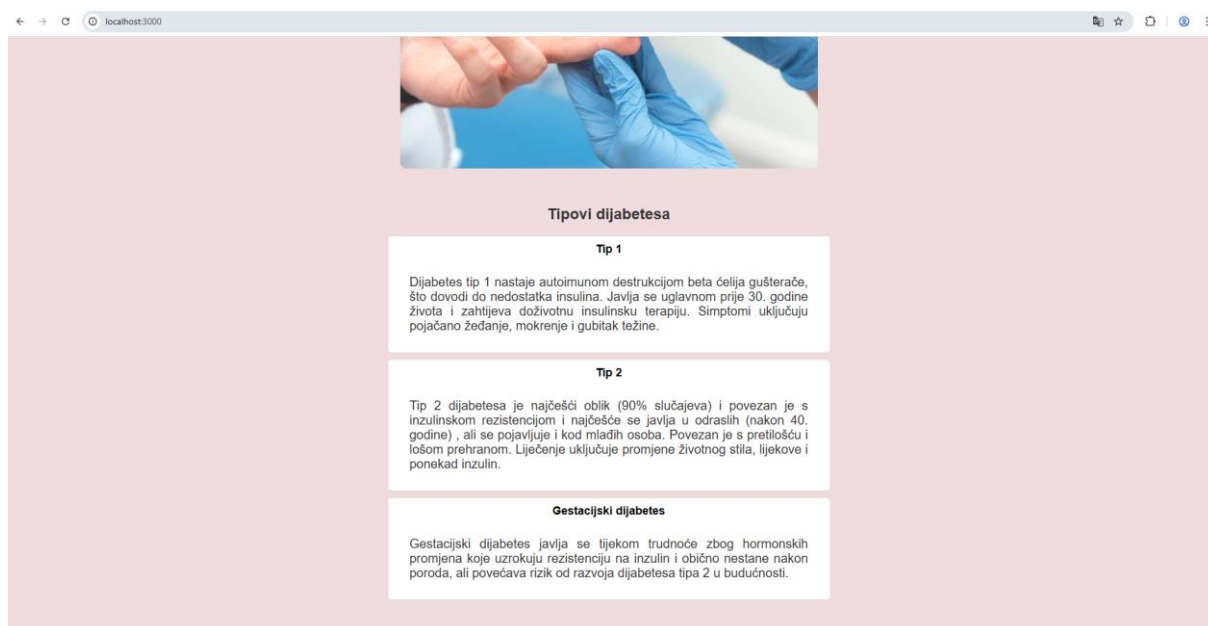
3.3. Klijentska aplikacija

Web stranice *Prediktor rizika od dijabetesa*, omogućuje korisnicima unos medicinskih i demografskih podataka kako bi se predvidio rizik od dijabetesa pomoću strojnog učenje modela implementiranog na *Azure Container Instances*. Na slici 26 prikazana je početna stranica. Korisnik se ovdje može informirati o tipovima dijabetesa koji postoje, podatci se mogu unijeti u formi za predikciju dijabetesa, a moguće je i informirati se o podacima koji su korišteni za izradu modela za predikciju.



Slika 26 Prikaz početne stranice

Na slici 27 dodatne informacije mogu se vidjeti, a one se prikazuju nakon pritiska na tipa šećerne bolesti koji nas zanima [8].



Slika 27 Prikaz dodatnih informacija za tipove dijabetesa

Na slici 28 prikazan je izgled forme, a rizik od dijabetes može se provjeriti unosom podatka o spolu, dobi, hipertenziji, bolesti srca, BMI-ju, razini hemoglobina A1c, razini glukoze u krvi i povijesti pušenja.

Slika 28 Prikaz forme

Na slici 29 prikazan je rezultat kada se prikazuje poruka „Postoji rizik od dijabetesa! Kontaktirajte liječnika što prije.“. U ovom primjeru osoba je ženskog spola od 30 godina koja nema problema s srcem i krvnim tlakom, koja ima normalni BMI, veliku prosječnu razinu glukoze u krvi tijekom 2-3 mjeseca (HbA1c) što je snažni pokazatelj postojanja dijabetesa, normalnu razinu glukoze u krvi, a bivši je pušač.

Slika 29 Prikaz rezultata za pacijenta koji ima veliki rizik za razvijanje dijabetesa

Slika 30 prikazuje drugi slučaj kada se prikazuje poruka „Nema indikacija dijabetesa. Nastavite zdravim životnim navikama“. U ovom primjeru riječ je o muškoj osobi od 28 godina koja nema problema s srcem i krvnim tlakom, koja ima blago povišen BMI, visoku razinu HbA1c i normalnu razinu glukoze u krvi, a trenutno puši.

Provjera rizika od dijabetesa

Spol: Muško

Dob: 28

Hipertenzija: Ne

Bolest srca: Ne

BMI (kg/m²): 27

HbA1c razina: 6.6

Glukoza u krvi (mg/dL): 80

Povijest pušenja: Trenutno pušim

Provjeri rizik

Napomena: Ovo nije medicinska dijagnoza!

Nema indikacija dijabetesa. Nastavite zdravim životnim navikama!

Analiza skupa podataka

Za izradu ovog modela za predviđanje rizika od dijabetesa koristio se skup podataka koji sadrži 100 000 primjeraka o pacijentima.

Slika 30 Prikaz rezultata za pacijenta koji ima mali rizik za razvijanje dijabetesa

Slika 31 prikazuje dodatni opis varijabli koje se moraju unijeti u ovom slučaju za varijablu BMI, gdje su navedene normalne i ekstremne vrijednosti ili rasponi vrijednosti. Dodatni opisi su dostupni za varijable hipertenzija, BMI, HbA1c i razinu glukoze u krvi.

Provjera rizika od dijabetesa

Spol: Muško

Dob:

Hipertenzija:

Bolest srca:

BMI (kg/m²): 27

HbA1c razina: 6.6

Glukoza u krvi (mg/dL): 80

Povijest pušenja: Trenutno pušim

Provjeri rizik

Napomena: Ovo nije medicinska dijagnoza!

Nema indikacija dijabetesa. Nastavite zdravim životnim navikama!

Analiza skupa podataka

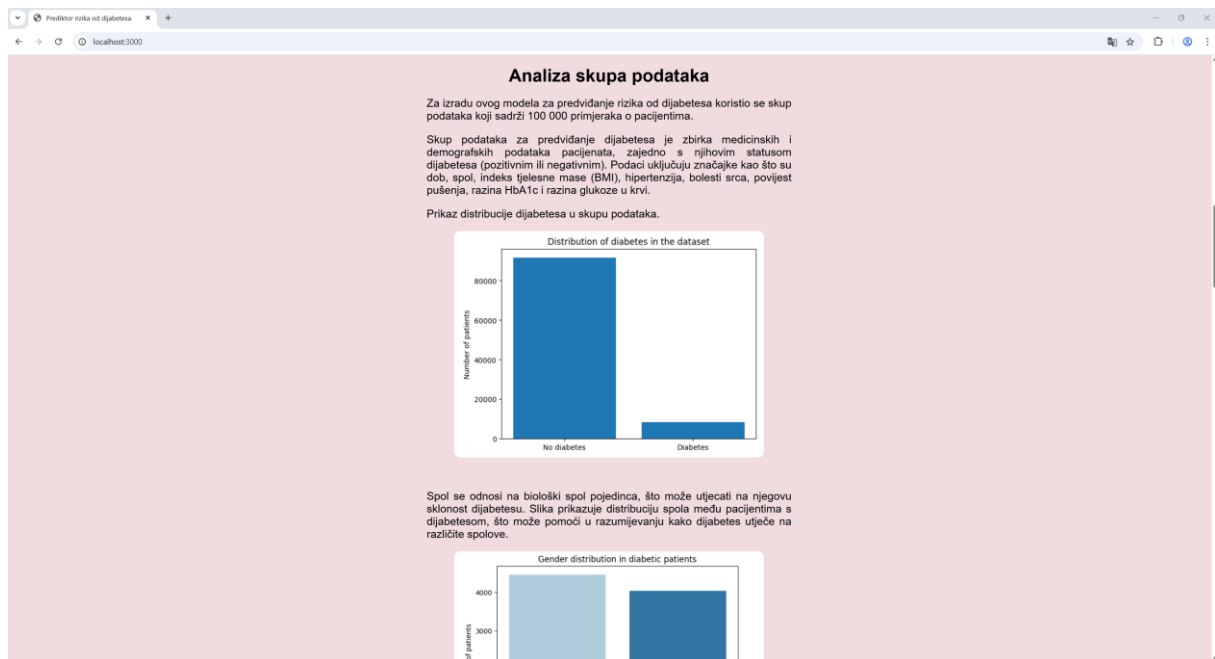
Za izradu ovog modela za predviđanje rizika od dijabetesa koristio se skup podataka koji sadrži 100 000 primjeraka o pacijentima.

BMI je mjera tjelesne mase na temelju težine i visine. Računa se kao visina(masa/masa). Mjera kao indikacija za raznog stupnja bolesti tipa 2, povišenog krvnog tlaka i kardiovaskularnih bolesti.

<18.5%	Podhranjenost
18.5-24.9%	Normalna težina
25-29.9%	Prekompjerna težina
>30%	Prestup

Slika 31 Prikaz dodatnih informacija

Ispod forme za predviđanje rizika od dijabetesa nalazi se dio koji je posvećen boljem shvaćanju podataka koji su korišteni za izradu ovog predikcijskog modela, a to je vidljivo na slici 32. U tom dijelu su prikazani dijagrami raspodjele podataka zajedno s opisom zašto se te varijable smatraju značajnima za predviđanje dijabetesa.



Slika 32 Prikaz informacija koje su korištene za izradu modela

4. Zaključak

Za realizaciju rješenja *Predviđanje dijabetesa* korišten je skup podataka od 100.000 pacijenata s 8 značajki : spol, dob, hipertenzija, srčane bolesti, BMI, HbA1c, glukoza u krvi i povijest pušenja. Podaci su analizirani (distribucije, korelacije prikazane na slikama 2-9) i obrađeni pomoću *one-hot encoding* tehnike za kategorijsku značajku *povijest pušenja* (6 kategorija: Nema informacija, trenutni pušač, pušio sam nekada, bivši pušač, nikada, ne pušim trenutno), *spol* je mapiran u binarnu vrijednost, te preuzorkovana je manjinske klase (dijabetes: 8.5%) na 30% omjera (slike 11-12).

Nakon što je logistička regresija odabrana kao optimalni algoritam, hiperparametri su optimizirani korištenjem *Azure Automated ML-a*, korištenjem 5-unakrsne validacije, a model je treniran u *Azure ML Studio* koristeći *Designer* cjevovod (slike 19, 21, 23) s 80/20 *train/test* podjelom. Ostvarene su zadovoljavajuće metrike (slika 22), adresa krajnje točke nalazi se na <http://761e0046-f8d3-4485-af0d-7ff6f8b1f918.polandcentral.azurecontainer.io/score>, gdje se prima POST JSON {Inputs: {data: [...]}, GlobalParameters: {method: "predict"}} i vraća {Results: [0/1]} (Tablica 1).

Razvijena je web aplikacija *Prediktor rizika od dijabetesa* s jednostavnom HTML formom (slika 28), na kojoj su smještene edukativne informacije za tipove dijabetesa (slika 27), kao i analiza podataka korištenih za izradu predikcijskog model (slika 32).

Izrađeni model i aplikacija ispunjavaju sve zahtjeve za ranu detekciju dijabetesa, nudeći brzu, pouzdanu predikciju i korisnički prilagođeno sučelje za pacijente i zdravstvene djelatnike.

5. Poveznice i literatura

Programskom je rješenju moguće pristupiti preko:

[Programsko rješenje na GitHubu](#)

- [1] M. Kardum Pejić, *Šećerna bolest ili dijabetes*, PLIVAzdravlje, Hrvatska, 2021. Dostupno na: <https://www.plivazdravlje.hr/tekst/clanak/16158/secerna-bolest-ili-dijabetes.html> (pristupljeno: 01.02.2026.)
- [2] Microsoft Azure, Microsoft, online cloud platforma, dostupno na: <https://azure.microsoft.com/en-us/> (pristupljeno: 31.01.2026.)
- [3] Mohammed Mustafa, Diabetes Prediction Dataset, Kaggle, online dataset, 2021. Dostupno na: <https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset> (pristupljeno: 19.01.2026.)
- [4] Project Jupyter, Jupyter.org, online platforma za interaktivni rad s notebook-ovima, dostupno na: <https://jupyter.org/> (pristupljeno: 19.01.2026.)
- [5] Koja je normalna razina šećera u krvi?, Glukoza.hr, online zdravstveni članak, 2023. Dostupno na: <https://glukoza.hr/koja-je-normalna-razina-secera-u-krvi/> (pristupljeno: 02.02.2026.)
- [6] BMI – Indeks tjelesne mase, PLIVAzdravlje, online zdravstveni članak, 2026. Dostupno na: <https://www.plivazdravlje.hr/zdravlje-online/bmi> (pristupljeno: 01.02.2026.)
- [7] S. Đuderija, Šta je HbA1c – test?, ABC Doctors, online zdravstveni članak, objavljeno 27.02.2021. Dostupno na: <https://www.abc-doctors.com/sta-je-hba1c-test> (pristupljeno: 02.02.2026.)
- [8] Dijabetes - koji oblici postoje?, Poliklinika Mazalin, Hrvatska, objavljeno 25.05.2023. Dostupno na: <https://poliklinika-mazalin.hr/blog/dijabetes-koji-oblici-postoje/> (pristupljeno: 02.02.2026.)