

Fakultet elektrotehnike, računarstva i informacijskih tehnologija Osijek

Računarstvo usluga i analiza podataka

SEMINARSKI RAD

Predviđanje korištenja kokaina na temelju podataka o osobnosti

Matej Džijan

Osijek, 2019.

Sadržaj

1. Uvod	1
2. Opis problema	2
2.1. Korišteni podaci	2
2.1.1. Predobrada podataka	4
2.2. Korišteni postupci strojnog učenja	5
3. Opis programskog rješenja	9
3.1. Model strojnog učenja	9
3.2. Način korištenja API-ja	11
3.3. Klijentska aplikacija	12
3.4. Osobno testiranje	16
4. Zaključak	17
5. Poveznice i literatura	18

1. Uvod

Problem evaluacije većeg rizika za korištenje droga je vrlo važan. Korištenje droga je rizično ponašanje koje se ne događa u izolaciji, važan je faktor za povećanje rizika od lošeg zdravlja, zajedno s bržim mortalitetom i morbiditetom, te ima značajne posljedice na društvo. Ima brojne faktore rizika koje se definiraju kao bilo koji atribut, karakteristika ili događaj u životu koji povećava šansu za korištenje droga. Većina faktora vezanih za prvo korištenje droga uključuju psihološke, društvene, individualne, ekonomske i faktore vezane za okolinu. Dobar dio ovih faktora su povezani s osobinama ličnosti. Iako su legalne supstance poput šećera, alkohola i duhana odgovorne za puno veći broj ranih smrti od ilegalnih droga, društvene i osobne posljedice rekreacijskih droga mogu biti vrlo problematične.

Kokain je psihoaktivna droga koja se najčešće koristi rekreacijski. Pojam *psihoaktivna droga* označava one droge koje imaju nekakvo djelovanje na mentalne funkcije. Djelovanje kokaina je dovoljno prijatno ili zanimljivo da ga ljudi koriste upravo iz tih razloga.

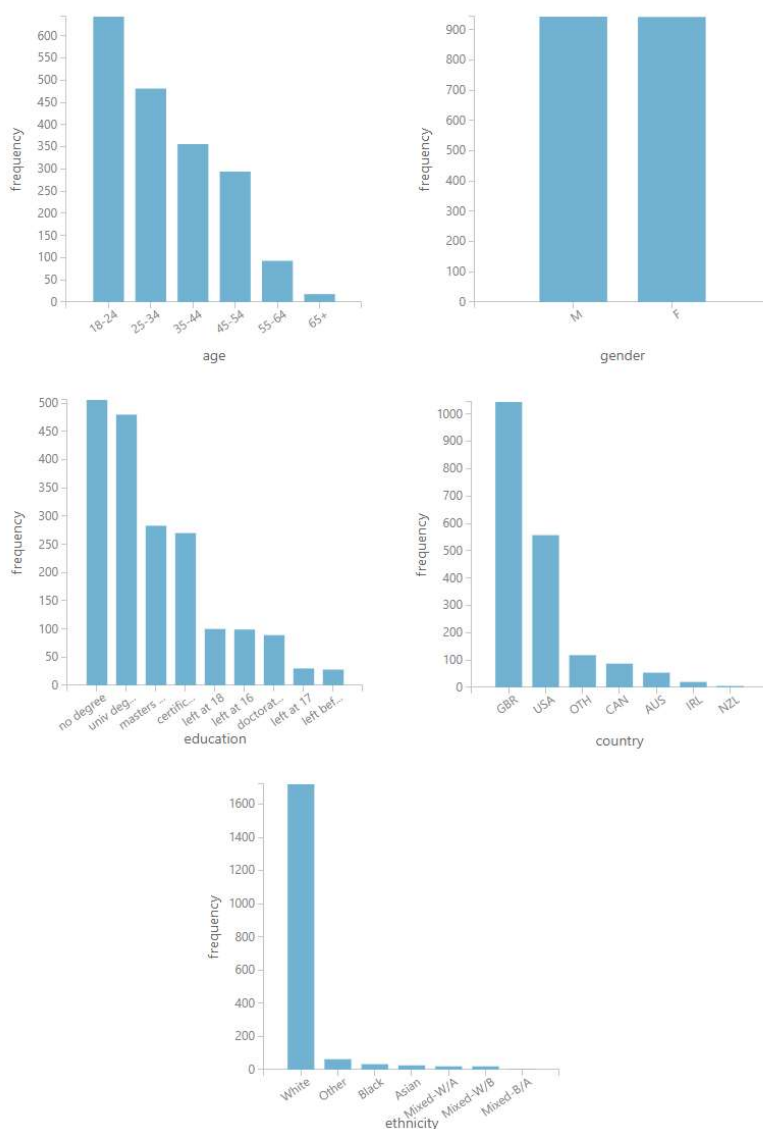
Cilj ovog projektnog zadatka je na primjeru kokaina predvidjeti je li osoba koristila kokain u zadnjih deset godina na temelju određenih osobnih parametara. Ti osobni parametri uključuju generalne podatke o osobi i sedam osobina ličnosti.

2. Opis problema

Ovaj zadatak je problem obične binarne klasifikacije. Za svaki skup ulaznih podataka, određeni parametri neke osobe, cilj je dobiti pretpostavku o tome je li ta osoba koristila kokain u zadnjih deset godina ili ne. Također, rezultati koju predviđaju visoku šansu da osoba koristi kokain, iako ga ne koristi, mogu ukazati na visok rizik za korištenje kokaina pa se može djelovati kako bi se spriječilo korištenje prije nego što se ono dogodi.

2.1. Korišteni podaci

Skup korištenih podataka se sastoji od podataka o 1885 sudionika istraživanja. Skup je u CSV formatu s 32 stupca: id, 12 stupaca osobnih podataka, 18 raznih droga te jednom izmišljenom drogom. Promatrani osobni podaci su sljedeći: dobna skupina, spol, edukacija, država, etnicitet, 5 NEO-FFI-R parametara (neuroticizam, ekstraverzija, otvorenost za iskustvo, susretljivost i savjesnost), BIS-11 (impulzivnost) i ImpSS (traženje osjeta – *sensation seeking*). Ovi osobni podaci su ulaz. Od ostalih podataka, korišteni su samo podaci o korištenju kokaina. Podatke su originalno prikupili Elaine Fehrman, Vincent Egan i Evgeny M. Mirkes.



Slika 2.1. Prikaz demografije sudionika istraživanja

Na slici 2.1. se vidi demografija sudionika istraživanja. Može se vidjeti da je raspodjela vrlo dobra za sve parametre osim za etnicitet i državu. Ovako mali broj pripadnika specifičnih ne-bijelih etničkih grupa isključuje svaku analizu koja uključuje rasne kategorije.

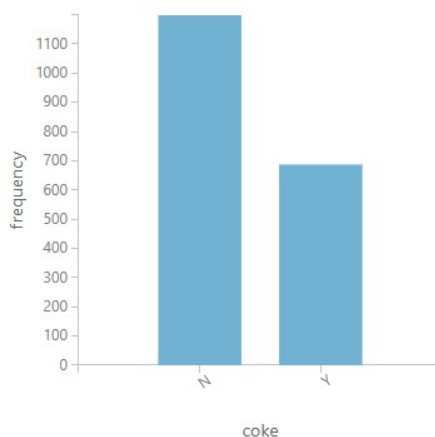


Slika 2.2. Raspodjela ostalih ulaznih atributa

Ostali ulazni atributi su osobine ličnosti (Slika 2.2.). Redom su to:

- Neuroticizam (Neuroticism) – dugoročna sklonost doživljavanja negativnih emocija kao što su nervoza, napetost, anksioznost i depresija
- Ekstraverzija (Extraversion) – očituje se u druženju, toplini, aktivnosti, asertivnosti, prikljivosti, veselju i u potrazi za stimulacijom
- Otvorenost za iskustvo (Openness to experience) – općenito cijenjenje umjetnosti, neobičnih ideja, te maštovitih, kreativnih, nekonvencionalnih i širokih interesa
- Susretljivost (Agreeableness) – dimenzija međuljudskih odnosa koju karakterizira altruizam, povjerenje, skromnost, ljubaznost, suosjećanje i kooperativnost
- Savjesnost (Conscientiousness) – tendencija za organiziranost, snagu volje, upornost, pouzdanost i učinkovitost
- Impulzivnost (Impulsiveness) – tendencija djelovanja na impuls, prikazuje ponašanje koje karakterizira malo ili nimalo promišljanja, razmišljanja ili razmatranja posljedica
- Traženje osjeta (Sensation seeking) – potraga za iskustvima i osjećajima koji su raznoliki, novi, složeni i intenzivni, te spremnost na fizičke, socijalne, pravne i financijske rizike radi takvih iskustava

Od izlaznih podataka su korišteni samo podaci o korištenju kokaina. Ti podaci su izvorno bili u 7 klasa: „Nikad korišteno“, „Korišteno prije više od 10 godina“, „Korišteno u zadnjih 10 godina“, „Korišteno u zadnjih godinu dana“, „Korišteno u zadnjih mjesec dana“, „Korišteno u zadnjih tjedan dana“ i „Korišteno u zadnja 24 sata“. Odlučio sam te podatke staviti u dvije klase, „Nije korisnik“ i „Korisnik“, gdje su prve dvije klase izvornog skupa, „Nikad korišteno“ i „Korišteno prije više od 10 godina“, bile svrstane pod „Nije korisnik“, a sve ostale pod „Korisnik“. Na slici 2.3. je prikazana raspodjela izlaznih podataka. 64% (1198) ispitanika spada u klasu „Nije korisnik“, a ostalih 36% (687) u klasu „Korisnik“.



Slika 2.3. Raspodjela izlaznih podataka

2.1.1. Predobrada podataka

Gore prikazane vizualizacije podataka su dobivene iz dodatno obrađenih podataka izvornog skupa.



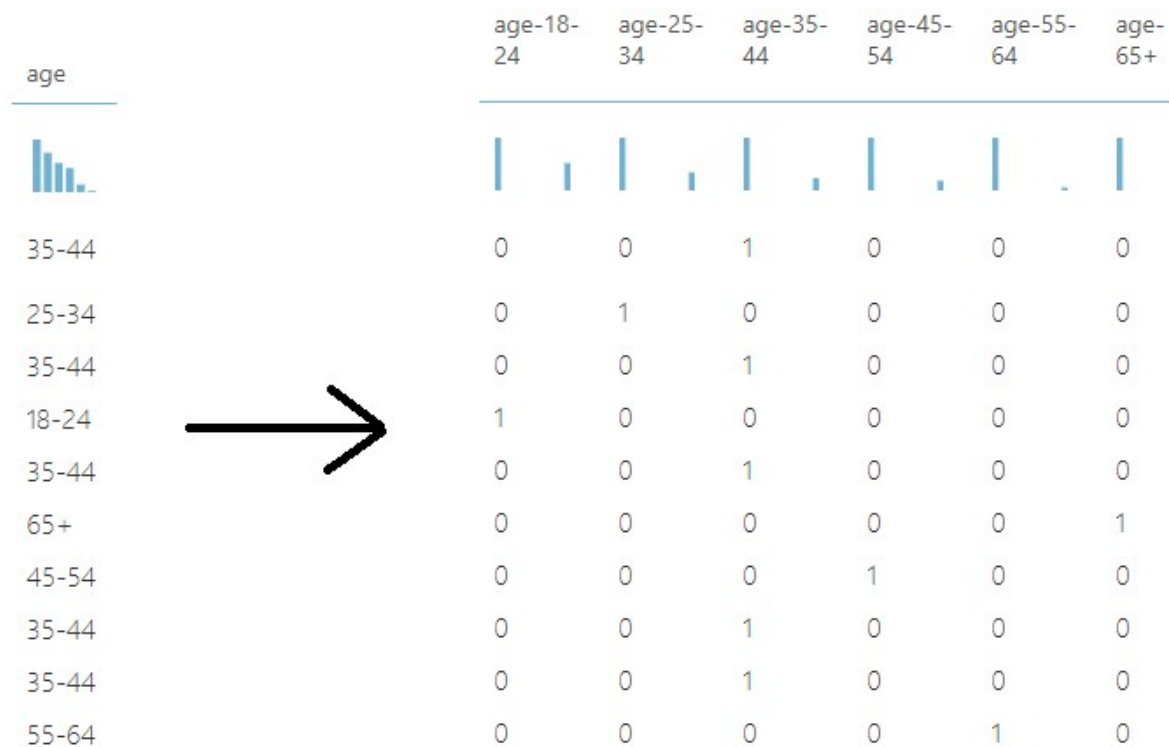
Slika 2.4. Predobrada podataka

Prvi korak je bio imenovanje stupaca kako bi se lakše snalazilo u podacima. Sljedeći korak je bio postavljanje izlaznih vrijednosti (podaci o korištenju droga) u kategoričke kako bi se mogli grupirati u dvije nove grupe, „Korisnik“ i „Nije korisnik“, odnosno „Y“ i „N“. Nakon toga su ti isti podaci grupirani u navedene grupe. Nakon toga je odabran samo stupac vezan za kokain kao „label“ jer je odlučeno raditi samo s tim stupcem. Nakon toga je pet atributa o osobnim podacima ispitanika (dobna skupina, spol, edukacija, država i etnicitet) prebačeno u oblik koji je prikazan na slici 2.1. jer su prethodno bili zapisani kao realni brojevi (Slika 2.5.).

```
1 import pandas as pd
2
3 def azureml_main(dataframe1 = None, dataframe2 = None):
4
5     dataframe1['age'] = dataframe1['age'].map({-0.95197 : '18-24', -0.07854 : '25-34', 0.49788 : '35-44', 1.09449 : '45-54', 1.82213 : '55-64', 2.59171 : '65+'})
6     dataframe1['gender'] = dataframe1['gender'].map({0.48246 : 'F', -0.48246 : 'M'})
7     dataframe1['education'] = dataframe1['education'].map({-2.43591 : 'left before 16', -1.7379 : 'left at 16', -1.43719 : 'left at 17', -1.22751 : 'left at 18',
8     -0.61113 : 'no degree', -0.05921 : 'certificate/diploma', 0.45468 : 'univ degree', 1.16365 : 'masters degree', 1.98437 : 'doctorate degree'})
9     dataframe1['country'] = dataframe1['country'].map({-0.09765 : 'AUS', 0.24923 : 'CAN', -0.46841 : 'NZL', -0.28519 : 'OTH', 0.21128 : 'IRL', 0.96082 : 'GBR', -0.57009 : 'USA'})
10    dataframe1['ethnicity'] = dataframe1['ethnicity'].map({-0.50212 : 'Asian', -1.10782 : 'Black', 1.90725 : 'Mixed-B/A', 0.126 : 'Mixed-W/A', -0.22166 : 'Mixed-W/B', 0.1144 : 'Other', -0.31685 : 'White'})
11
12
13
14
15    return dataframe1
```

Slika 2.5. Predobrada osobnih podataka iz realnih brojeva u stringove

Zatim su te vrijednosti označene kao kategoričke i, nakon toga, prebačene u indikatorske vrijednosti te su dobiveni rezultati kao što je prikazano na slici 2.6. za primjer dobi. Za svaku od kategorija se pravi novi stupac te se stavlja vrijednost 1 u stupac kojoj kategoriji pripada taj unos, a 0 u sve druge stupce.



Slika 2.6. Prebacivanje kategoričkog atributa dobi u indikatorske vrijednosti za svaku dobnu skupinu

Nakon ovoga se odabiru stupci koji će se koristiti pri strojnom učenju. Prvo se izbacuju stupci id, stari kategorički stupci za osobne podatke, te sve droge osim kokaina. Na posljetku se još vrši odabir značajki na temelju filtra (*Filter Based Feature Selection*) gdje se odabire manji broj stupaca za koje filter procijeni da su najpogodniji za predviđanje klase za stupac kokain. Odabrano je 15 značajki i sam stupac kokain zbog velikog pada procjene utjecaja između 15. i 16. značajke, a te značajke su: *ss*, *country-GBR*, *impulsive*, *country-USA*, *oscore*, *cscore*, *age-45-54*, *ascore*, *gender-F*, *gender-M*, *education-no degree*, *age-18-24*, *nscore*, *age-25-34*, *age-55-64* i *education-masters degree*.

Algorithm	Accuracy	Precision	Recall	F-Score	AUC	Average Log Loss	Training Log Loss
Feature Selection + Decision Jungle	0.732484	0.651515	0.518072	0.577181	0.771223	0.54187	16.499897
Decision Jungle	0.723992	0.630435	0.524096	0.572368	0.778432	0.532214	17.98785

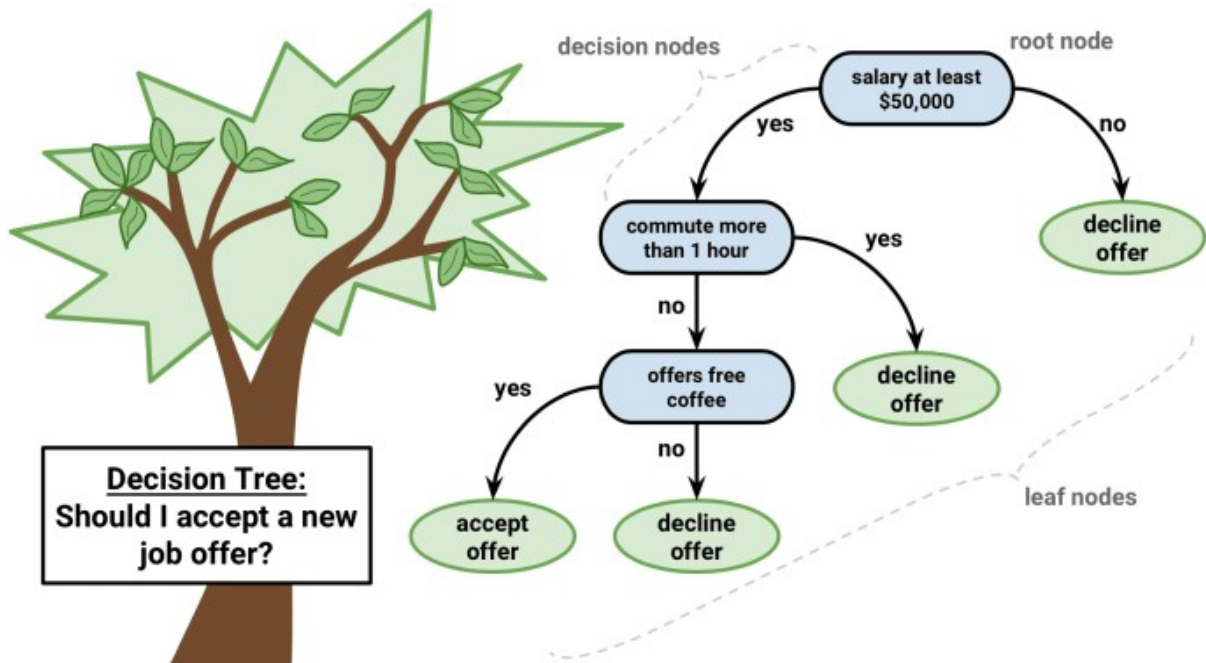
Slika 2.7. Usporedba rezultata istog algoritma s i bez odabira značajki na temelju filtra

2.2. Korišteni postupci strojnog učenja

Problem binarne klasifikacije se može riješiti na velik broj načina. Za ovaj zadatak su uspoređeni rezultati raznih postupci strojnog učenja, te je odabran onaj s najboljim rezultatima. Korišteni algoritmi strojnog učenja su: džungla odluke, šuma odluke, pojačano stablo odluke, lokalno-duboki stroj s vektorima podrške, stroj s vektorima podrške, logistička regresija i neuronska mreža. Za svaki od algoritama se koristilo automatsko podešavanje parametara kako bi se dobili najbolji parametri za svaki od algoritama, te su ti rezultati uspoređivani pri odabiru najboljeg algoritma.

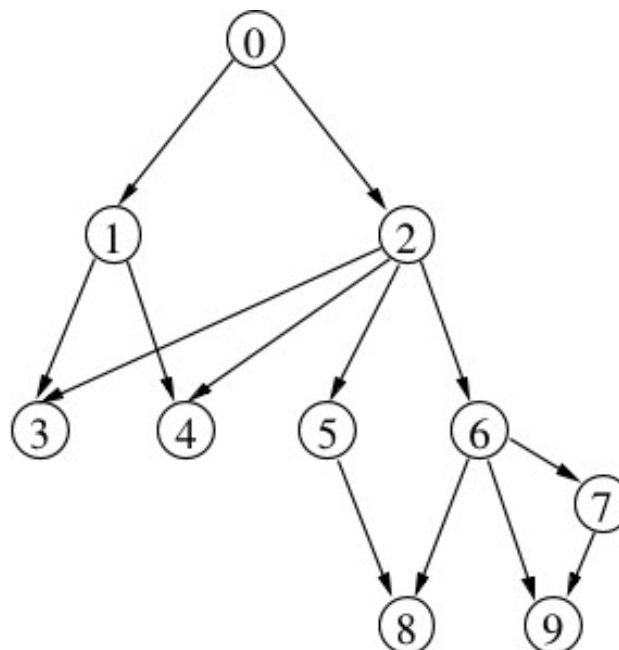
Stablo odluke je vrlo jednostavan algoritam koji radi na sljedeći način:

1. Postavlja se najbolji atribut skupa podataka kao korijen stabla
2. Skup podataka se dijeli u podskupove tako da svi elementi podskupa imaju jednaku vrijednost odabranog atributa
3. Ponavljanje koraka 1. i 2. na svakom podskupu dok se ne dođe do kraja na svakoj od grana drveta



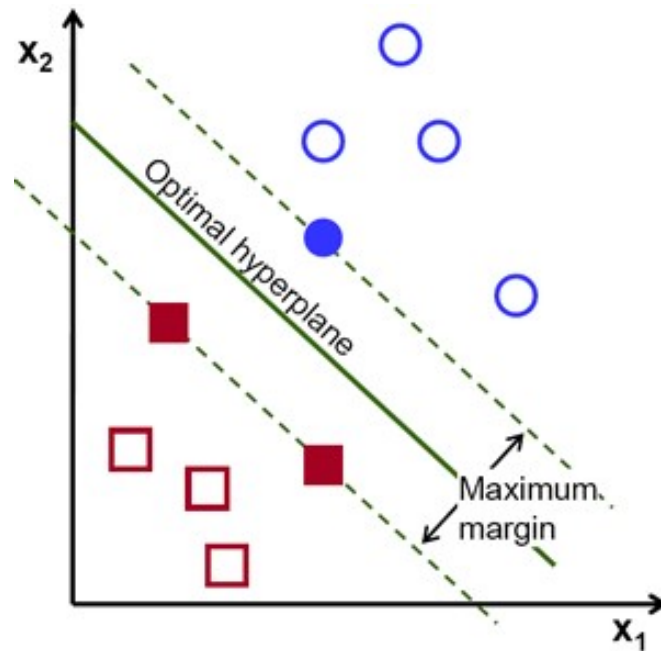
Slika 2.8. Primjer stabla odluke

Šume odluke su skup različitih stabala odluke, a konačni rezultat se dobije odabirom većinskog odgovora. Džungle odluke su nalik na šume odluke, no umjesto stabala, koriste se usmjereni aciklički grafovi (Slika 2.9.), gdje čvor može imati više od jednog roditeljskog čvora.



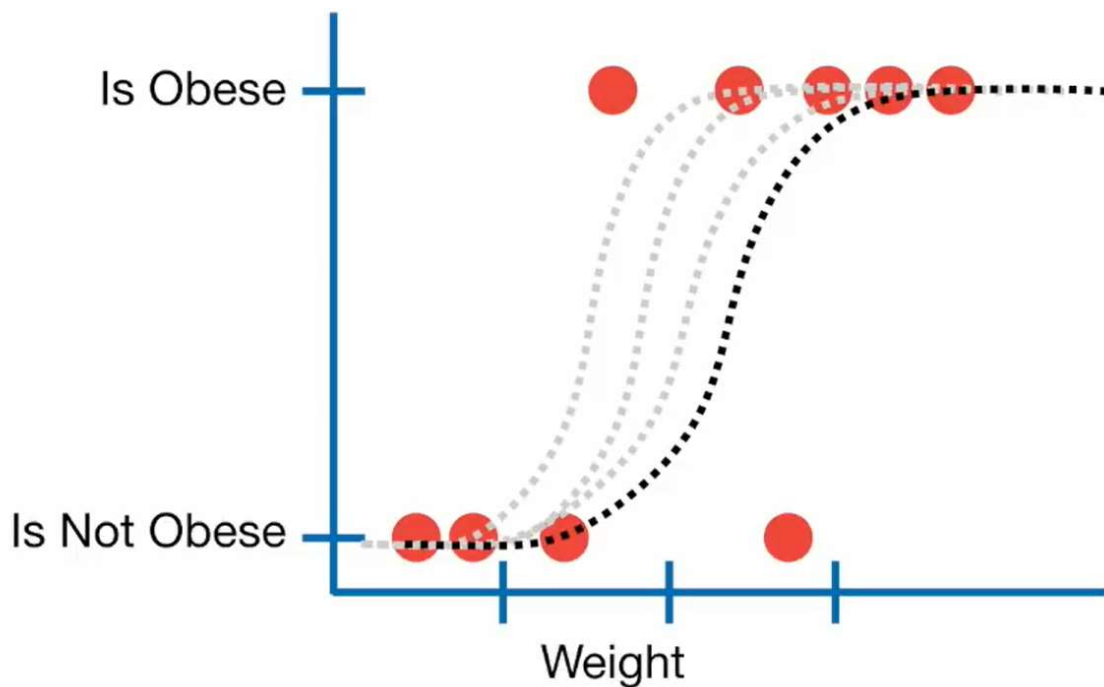
Slika 2.9. Usmjereni aciklički graf

Stroj s vektorima podrške konstruira hiper-ravninu ili set parametara u visoko-dimenzionalnom prostoru. Dobro odvajanje klasa se dobije hiper-ravninom koja ima najveću udaljenost do najbližeg podatka za treniranje bilo koje klase. Na slici 2.10. je prikazan primjer načina rada jednostavnog stroja s vektorima podrške.



Slika 2.10. Prikaz načina rada jednostavnog stroja s vektorima podrške

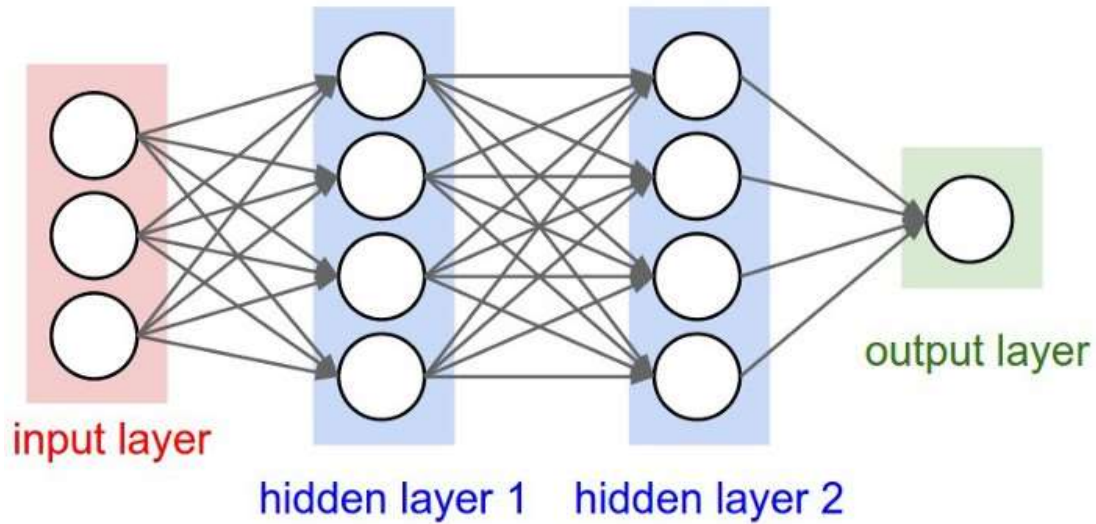
Logistička regresija je algoritam koji, ukratko, pokušava smjestiti krivulju koja će najbolje predvidjeti izlaznu klasu. Na slici 2.11. je prikazano nekoliko krivulja za predviđanje klase pretilosti na temelju mase. Algoritam računa koliko dobro određena krivulja predviđa klasu, te izabire najbolju krivulju.



Slika 2.11. Logistička regresija

Neuronska mreža je sustav neurona. Neuroni su organizirani u tri sloja:

1. Ulazni sloj – unosi podatke u sustav za daljnju obradu
2. Skriveni sloj – sloj između ulaznog i izlaznog gdje neuronska mreža uzima set ponderiranih ulaza i proizvodi izlaz kroz aktivacijsku funkciju
3. Izlazni sloj – zadnji sloj neurona koji daje izlaze za program



Slika 2.12. Prikaz neuronske mreže









Neuronska mreža uzima ulazni podatak, a zatim na temelju težina sinapsi između neurona i raznih drugih parametara računa izlazne vjerojatnosti za klase zadanog atributa. Nakon toga se težine namještaju i postupak se ponavlja za ostale ulazne podatke.

3. Opis programskog rješenja

Konačno rješenje uz klijentsku aplikaciju je zamišljeno za upotrebu psihologa kako bi predvidjeli potencijalan rizik pacijenta za upotrebu kokaina. Rješenje bi se dalje moglo prenijeti i na druge tipove droga.

3.1. Model strojnog učenja

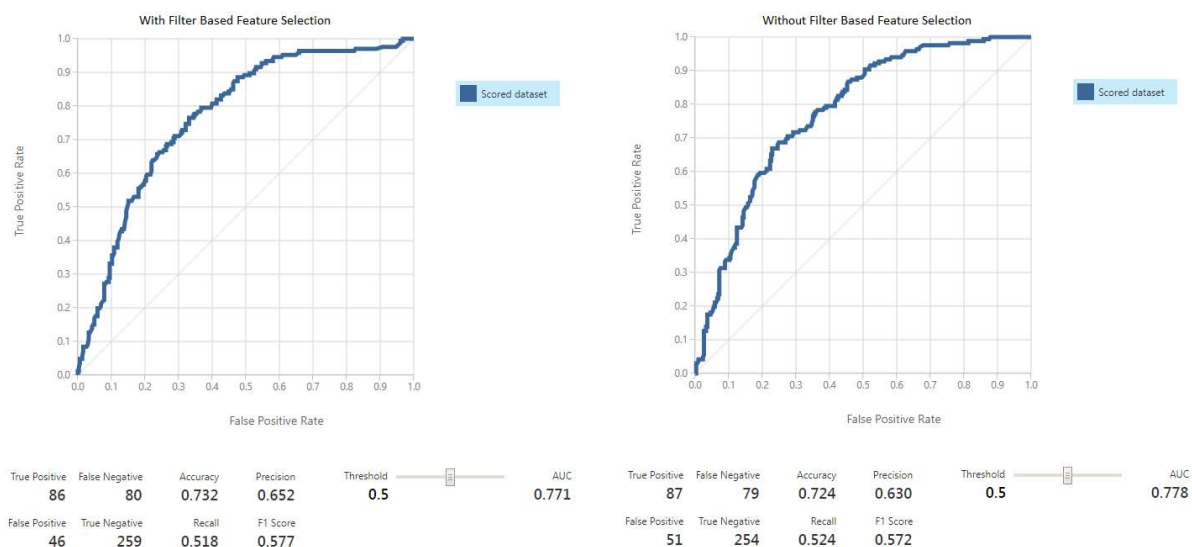
Kao što je prethodno navedeno, za klasifikaciju je isprobano 7 algoritama strojnog učenja, te su korišteni razni hiperparametri za svaki od njih. Prije treniranja modela, skup podataka je podijeljen na skup za treniranje i na skup za testiranje u omjeru 0.75 : 0.25. Svi modeli su testirani na skupu za testiranje te su dobiveni sljedeći rezultati:

Algorithm	Accuracy	Precision	Recall	F-Score	AUC	Average Log Loss	Training Log Loss
							
Decision Jungle	0.723992	0.630435	0.524096	0.572368	0.778432	0.532214	17.98785
Decision Forest	0.711253	0.615385	0.481928	0.540541	0.759648	0.549979	15.250253
Boosted Decision Tree	0.70276	0.590278	0.512048	0.548387	0.732175	1.547052	-138.394861
Locally-Deep SVM	0.694268	0.577465	0.493976	0.532468	0.684318	1.067132	-64.441045
Logistic Regression	0.723992	0.625	0.542169	0.580645	0.766186	0.5449	16.032951
Neural Network	0.715499	0.615942	0.512048	0.559211	0.764882	0.539173	16.915498
SVM	0.685775	0.569231	0.445783	0.5	0.752183	0.557917	14.027

Slika 3.1. Rezultati različitih algoritama

Iz slike 3.1. može se vidjeti da logistička regresija i džungla odluke imaju najvišu točnost, oko 72.4%, ali je odlučeno koristiti džunglu odluke. Iako je logistička regresija jednostavnija i tako poželjna ako su rezultati jednaki, izabrano je koristiti džunglu odluke zbog nižeg porsječnog *log lossa* koji je bitan u ovom slučaju gdje se interpretiraju i same vjerojatnosti, ne samo rezultatna klasa.

Nakon odabira modela, dodatno su unaprijeđeni performansi modela dodavanjem odabira značajki na temelju filtra, kao što je prikazano na slici 2.7. prethodno u radu, no na slici 3.2. je prikazana detaljna usporedba rezultata uz ROC funkciju. Poboljšanja nisu velika, ali su prisutna, iako se opoziv povećao vrlo malo.



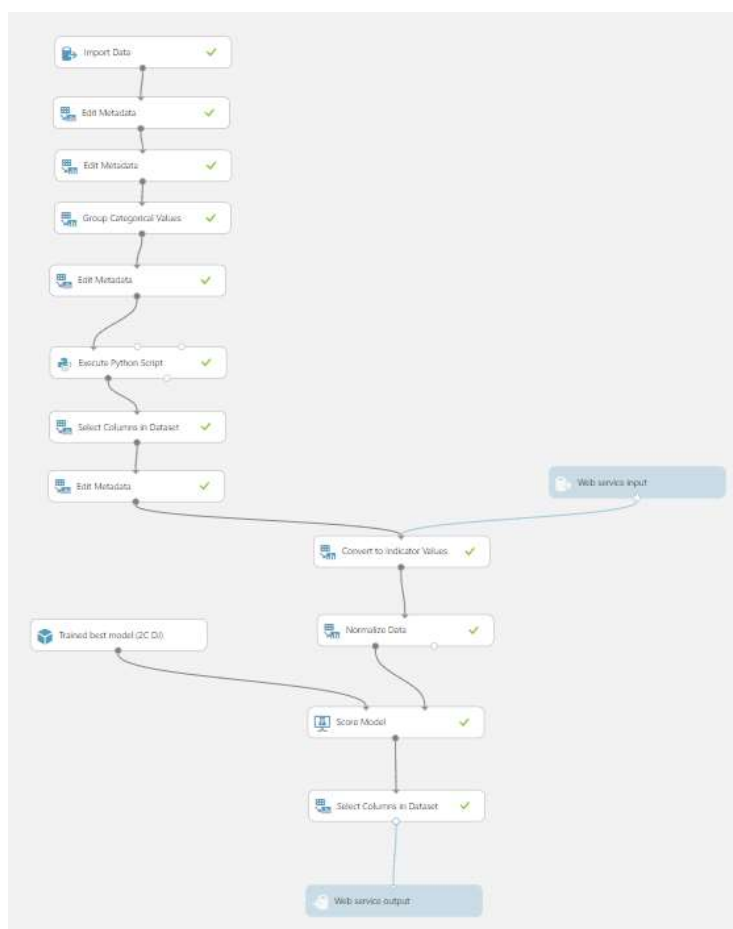
Slika 3.2. Detaljna usporedba modela s i bez odabira značajki na temelju filtra

Svaki od modela je testiran s brojem kombinacijama hiperparametara, na slici 3.3. je prikazana usporedba raznih parametara za odabrani model džungle odluke nakon odabira značajki na temelju filtra. Na vrhu je prikazana kombinacija najboljih parametara.

Number of optimization steps per decision DAG layer	Maximum width of the decision DAGs	Maximum depth of the decision DAGs	Number of decision DAGs	Accuracy	Precision	Recall	F-Score	AUC	Average Log Loss	Training Log Loss
8782	733	8	9	0.732484	0.651515	0.518072	0.577181	0.771223	0.54187	16.499897
3279	227	26	23	0.717622	0.618705	0.518072	0.563934	0.761229	0.545468	15.945379
10621	224	57	28	0.713376	0.616541	0.493976	0.548495	0.758157	0.616264	5.036042
12222	735	8	29	0.713376	0.624	0.46988	0.536082	0.764586	0.545659	15.915933
10544	500	13	28	0.711253	0.619048	0.46988	0.534247	0.760221	0.547921	15.567421
14957	814	22	15	0.707006	0.598592	0.512048	0.551948	0.755678	0.557639	14.069877
2353	858	11	20	0.704883	0.59854	0.493976	0.541254	0.75321	0.549343	15.348302
13702	753	55	22	0.692144	0.573427	0.493976	0.530744	0.732945	0.70425	-8.522223
13723	225	63	15	0.690021	0.568493	0.5	0.532051	0.740263	0.794857	-22.484498
11041	737	51	4	0.683652	0.571429	0.409639	0.477193	0.69524	1.160553	-78.836911

Slika 3.3. Razne kombinacije hiperparametara za odabrani model

Nakon odabira najboljeg modela, model je postavljen kao web usluga na Azureu. Podaci prolaze kroz dio pretprocesiranja kao i podaci na kojima je treniran model. Kategorijski podaci o osobi se prebacuju u indikatorske vrijednosti. Na posljertku se aplikaciji vraća samo vrijednost „Y“ ili „N“ je li osoba korisnik, te vjerojatnost (0-1).



Slika 3.4. Realizacija web servisa

3.2. Način korištenja API-ja

Korišteni API je jednostavan API koji Azure predlaže. Podaci se šalju u sljedećem obliku:

```
1. {
2.   "Inputs": {
3.     "input1": {
4.       "ColumnNames": [
5.         "age",
6.         "gender",
7.         "education",
8.         "country",
9.         "ethnicity",
10.        "nscore",
11.        "escore",
12.        "oscore",
13.        "ascore",
14.        "cscore",
15.        "impulsive",
16.        "ss",
17.        "coke"
18.      ],
19.      "Values": [
20.        [
21.          "18-24",
22.          "F",
23.          "certificate/diploma",
24.          "AUS",
25.          "Asian",
26.          "0",
27.          "0",
28.          "0",
29.          "0",
30.          "0",
31.          "0",
32.          "0",
33.          "N"
34.        ]
35.      ]
36.    }
37.  },
38.  "GlobalParameters": {}
39. }
```

Web servisu se šalju ulazni parametri kao što je navedeno. Podaci za dob, spol, edukaciju, državu i etnicitet se šalju kao stringovi, koji su na web servisu kategorije, a zatim se pretvaraju u indikatorske vrijednosti. Ostale ulazne vrijednosti se šalju kao realni brojevi. Također je potrebno poslati i neku vrijednost za parametar „coke“ koja neće utjecati na krajnji rezultat. Web servis šalje odgovor, koji se dalje interpretira u programu, u sljedećem obliku:

```
1. {
2.   "Results": {
3.     "output1": {
4.       "type": "DataTable",
5.       "value": {
6.         "ColumnNames": [
7.           "Scored Labels",
8.           "Scored Probabilities"
9.         ],
10.        "ColumnTypes": [
11.          "Categorical",
12.          "Numeric"
13.        ],

```

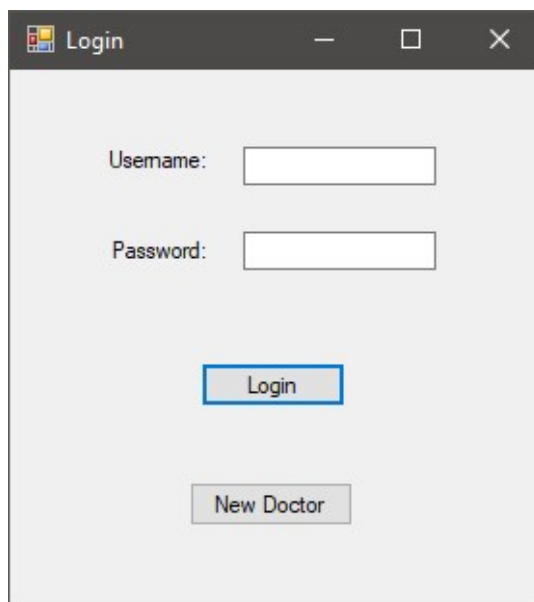
```

14.         "Values": [
15.             [
16.                 "N",
17.                 "0"
18.             ]
19.         ]
20.     }
21. }
22. }
23. }

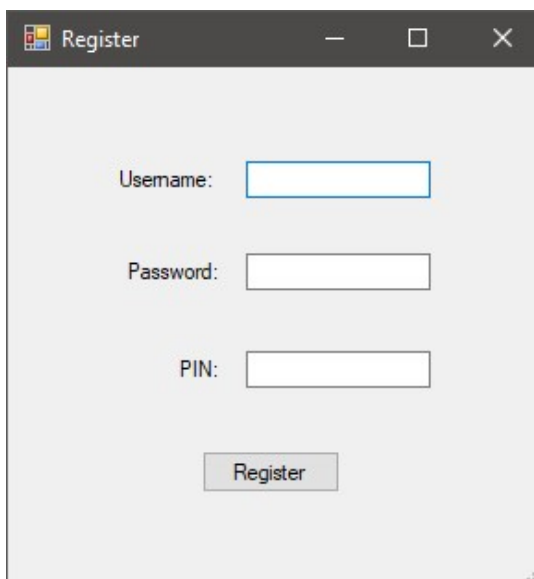
```

3.3. Klijentska aplikacija

Klijentska aplikacija je napravljena u obliku Windows Form aplikacije. Osmišljena je na način da se korisnik (doktor) prvo treba logirati u aplikaciju kako bi ju koristio. Ako doktor nema račun, može napraviti novi. Pravljenje računa je napravljeno tako da korisnik mora znati tajni PIN koji je ugrađen u kod. U nastavku su prikazani izgledi *login* (Slika 3.5.) i *register* (Slika 3.6.) formi.



Slika 3.5. *Login* forma



Slika 3.6. *Register* forma

	Id	Username	Password
	1	Mike	mik
	2	John	jif
+	NULL	NULL	NULL

Slika 3.7. Podaci o doktorima u bazi

Svi korisnici se spremaju lokalno korištenjem baza podataka. Za pristupanje bazi podataka korišten je EntityFramework NuGet paket. Pri loginu se provjerava ispravnost unosa (Slika 3.8.), te se otvara glavni prozor i zatvara trenutni ukoliko je unos ispravan.

```
private void buttonLogin_Click(object sender, EventArgs e)
{
    var username = textBoxUsername.Text;
    var password = textBoxPassword.Text;
    try
    {
        Models.Doctor doctor = dbContext.Doctors.Where(x => x.Username == username && x.Password == password).First();

        doctorId = doctor.Id;
        this.Hide();

        FormMain formMain = new FormMain();
        formMain.Show();
    }
    catch
    {
        MessageBox.Show("Wrong username or password.");
    }
}
```

Slika 3.8. Kod gumba za logiranje

Nakon što se korisnik prijavi, otvara se glavni prozor aplikacije (Slika 3.9.). U ovom prozoru je moguće unijeti podatke o novom pacijentu i pogledati statistike o prethodnim unosima o pacijentima.

Slika 3.9. Glavni prozor aplikacije

Nakon što se postave svi parametri, potrebno je pritisnuti gumb „Predict Cocaine Use Risk“ koji će zatim u *labelu* ispod ispisati rizik za unesenog pacijenta. Moguće je ispisati 6 različitih tekstova ovisno o vjerojatnosti ishoda koje web servis vraća kao što je prikazano na slici 3.10.

```
private void EditResultText()
{
    if (predictionResult > 0.8) labelResult.Text = "Patient has extremely high chance of being a user or becoming one. " +
        "I hope you didn't need this app to tell you that.";
    else if (predictionResult > 0.6) labelResult.Text = "There's a very high chance the patient is a user or has a very high risk of becoming one.";
    else if (predictionResult > 0.5) labelResult.Text = "This patient has decent chance for cocaine use.";
    else if (predictionResult > 0.4) labelResult.Text = "Patient probably doesn't use, but there is some risk of becoming a user.";
    else if (predictionResult > 0.2) labelResult.Text = "Patient has really low chance for cocaine use. There's no need for concern";
    else labelResult.Text = "Patient has extremely low chance for cocaine use.";
}
```

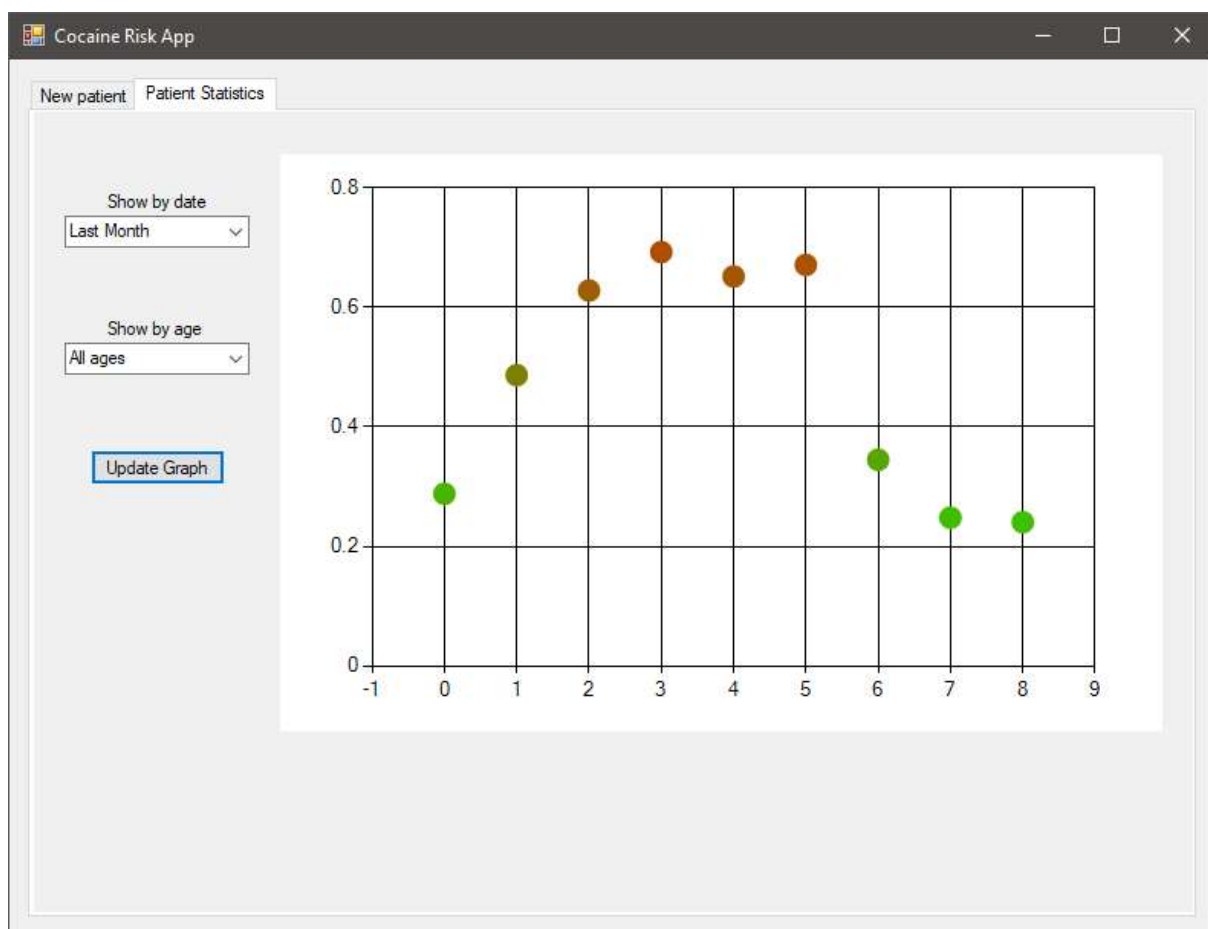
Slika 3.10. Mogući ispisi rezultata

Svaki od parametara daje jednostavan opis pri *hoveru*, a pri namještanju parametara, pojavljuje se *tooltip* koji omogućuje precizno namještanje parametara (Slika 3.11.).

Slika 3.11. *Tooltip* za lakši odabir vrijednosti parametara

Nakon što se predvidi rizik za unesenog pacijenta, taj pacijent se sprema u bazu zajedno sa svim informacijama, ali i s doktorom koji ga je unio tako da doktori ne mogu pristupati pacijentima drugih doktora.

Korisniku je također omogućen pregled prethodnih unosa i iscrtavanje na grafu (Slika 3.12). Za pregled pacijenata omogućen je odabir vremena unosa pacijenta u 5 kategorija: posljednjih mjesec dana, posljednjih 6 mjeseci, posljednjih godinu dana, posljednjih 5 godina i cijela povijest. Također je omogućen odabir dobne skupine koja se prikazuje.



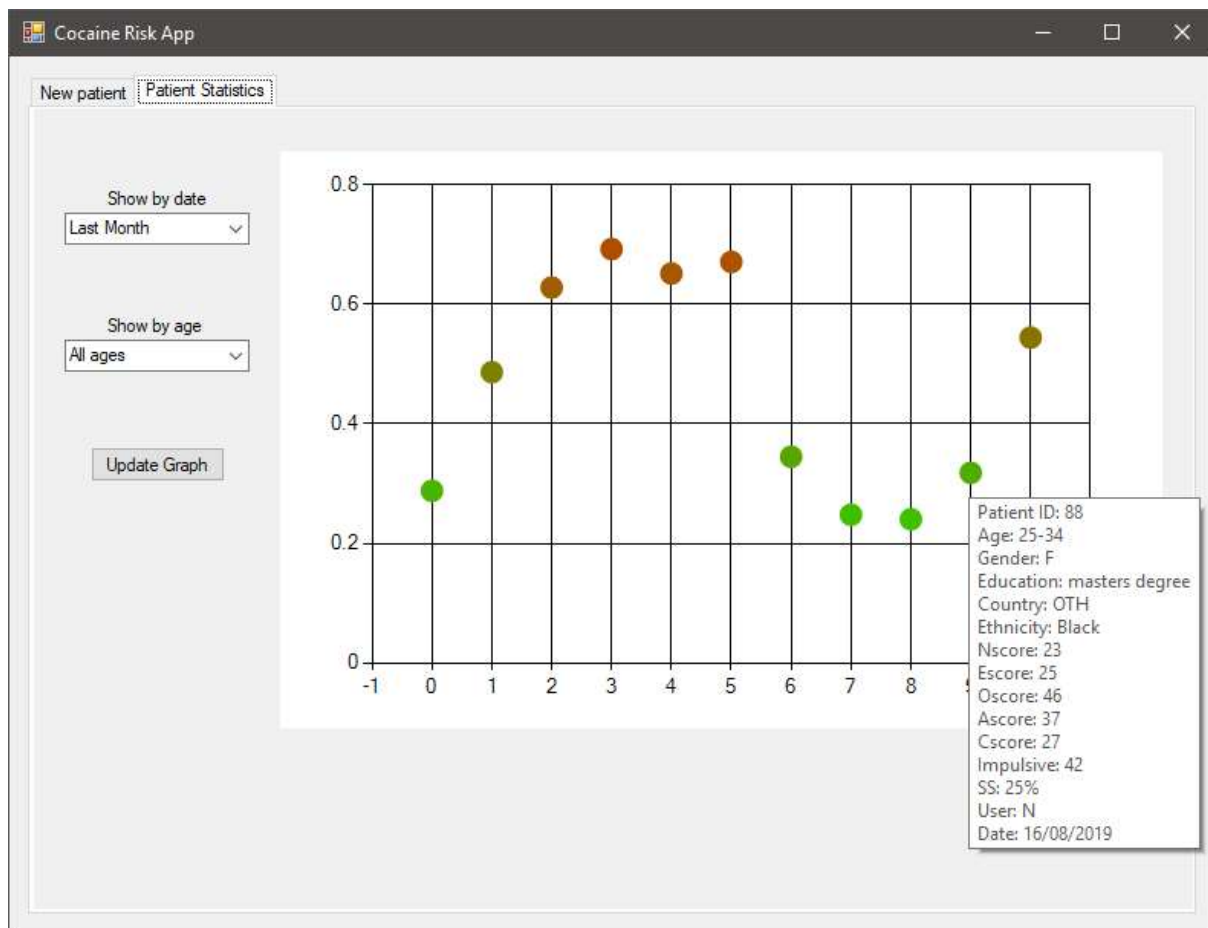
Slika 3.12. Iscrtavanje rezultata pacijenata na grafu

```
private void UpdateGraph()
{
    chartResults.Series["Result"].Points.Clear();
    results.Clear();
    int i = 0;
    int months;
    switch (comboBoxPeriod.SelectedIndex)
    {
        case (0):
            months = 1;
            break;
        case (1):
            months = 6;
            break;
        case (2):
            months = 12;
            break;
        case (3):
            months = 60;
            break;
        default:
            months = 0;
            break;
    }
    DateTime date;
    if (months != 0) date = DateTime.Now.AddMonths(-months);
    else date = DateTime.MinValue;
    var age = comboBoxShowByAge.Text;
    List<Models.Patient> patients;
    if (age == "All ages") patients = dbcontext.Patients.Where(x => x.Date >= date && x.DoctorId == doctorId).ToList();
    else patients = dbcontext.Patients.Where(x => x.Date >= date && x.DoctorId == doctorId && x.Age == age).ToList();

    foreach (var a in patients)
    {
        results.Add(a.ToPatient());
        chartResults.Series["Result"].Points.AddXY(i, a.Coke);
        chartResults.Series["Result"].Points[i].Color = Color.FromArgb((int)(a.Coke * 255), (int)((1 - a.Coke) * 255), 0);
        ++i;
    }
}
```

Slika 3.13. Dohvaćanje pacijenata iz baze podataka

Osim samog iscrtavanja, moguće je staviti pokazivač na određenog pacijenta, što će zatim prikazati detalje o tom pacijentu kao što je prikazano na slici 3.14. Detalji koji se prikazuju su svi uneseni podaci o pacijentu, njegov ID, rezultat te datum unosa pacijenta.



Slika 3.13. Detalji o prethodno unesenom pacijentu

3.4. Osobno testiranje

Nakon izrade, aplikacija je testirana na stvarnim ljudima. Nažalost, nijedan od ispitanika nije koristio kokain u posljednjih 10 godina i svi su dio iste dobne skupine (18-24). Različiti psihološki parametri su prikupljeni korištenjem online alata. Model predviđa da 1 od 4 ispitane osobe je koristila kokain u posljednjih 10 godina, što je pogreška. Uz preciznost modela od 73,25%, rezultati (75%) na ovako malom skupu su vrlo blizu očekivanih.

Tablica 3.1. Rezultati testiranja

Age	Gender	Education	Country	Ethnicity	Nscore	Escore	Oscore	Ascore	Cscore	Impulsive	Ss	Result
18-24	M	univ degree	OTH	White	28	47	54	25	21	78	73%	0.5966
18-24	F	univ degree	OTH	White	30	19	53	47	37	71	63%	0.484
18-24	M	univ degree	OTH	White	31	13	26	43	32	68	30%	0.2924
18-24	M	no degree	OTH	White	31	43	47	34	19	78	43%	0.4033

4. Zaključak

Predviđanje ponašanja ljudi je vrlo kompleksan problem. Ljudske osobine se ne mogu lagano kvantificirati, ali određeni psihološki testovi donekle omogućuju upravo to. Ta kvantifikacija ljudskih osobina je u ovom projektu korištena za predviđanje ljudskog ponašanja, odnosno korištenje kokaina. Korištenjem algoritma džungle odluka, postignuta je točnost od 73.25% što je vrlo blizu rezultata koje su postigli znanstvenici koji su izvorno prikupili korišteni skup podataka. Upravo zbog kompleksnosti ljudi, vjerujem da ovaj rezultat nije moguće znatno poboljšati bez dodatnih testova i parametara. Postupci slični onima koji su korišteni u ovom projektu se mogu primijeniti i na razne druge droge ili skupine droga pa tako predvidjeti potencijalne rizike i spriječiti neželjeno ponašanje pacijenata.

5. Poveznice i literatura

Programskom je rješenju moguće pristupiti preko:

[Programsko rješenje na GitHubu](#)

[Azure eksperiment](#)

[Azure ML model](#)

- [1] <https://arxiv.org/pdf/1506.06297.pdf>
- [2] <https://archive.ics.uci.edu/ml/datasets/Drug+consumption+%28quantified%29>
- [3] <http://psychtests.co.nf/sensation/test.html>
- [4] <https://www.psytoolkit.org/survey-library/impulsiveness-barratt.html>
- [5] <https://www.truity.com/test/big-five-personality-test>