



**УНИВЕРЗИТЕТ „СВ. КЛИМЕНТ ОХРИДСКИ“ – БИТОЛА
ТЕХНОЛОШКО ТЕХНИЧКИ ФАКУЛТЕТ ВЕЛЕС**



Иновативни технологии за храна и нутриционизам

**ПРИМЕНЕНА НА МАТЕМАТИЧКИ МОДЕЛИ И ТЕХНИКИ ЗА
ПРЕДИКЦИЈА НА ДИЈАБЕТЕС ТИП 2**
докторски проект

Кандидат
Валбона Мазлами

Ментор
ред.проф.д-р Весна Антоска Knights

ПРИМЕНЕНА НА МАТЕМАТИЧКИ МОДЕЛИ И ТЕХНИКИ ЗА ПРЕДИКЦИЈА НА ДИЈАБЕТЕС ТИП 2

Валбона Мазлами

Универзитет “Св. Климент Охридски” – Битола, Р.С. Македонија

ORCID iD

valbona.mazlami@uklo.edu.mk

Весна (Антоска) Knights

Универзитет “Св. Климент Охридски” – Битола, Р.С. Македонија

0000-0003-1768-2231

vesna.knights@uklo.edu.mk

Апстракт

Дебелината станува се поголем проблем на дешницата, проследена со други болести односно нарушување на општото здравје на човекот. Овај доктоски проект ќе се осврне од аспект на математичка анализа на клинички здравствени податоци кај 100 пациенти со дебелината, со цел на предикција и дијабетес мелитус тип 2 (T2DM). Податоците се прибирани од Клиничката болница „Мајка Тереза“ во Скопје во текот на 9 месеци. Множеството на податоци вклучува: број на пациент (поради заштита на податоци), тежина, пол, возраст, датум на контрола, БМИ, статус на контрола на тежина, дијагноза на дијабетес, класификација на дебелина, холестерол (mmol/l), триглицериди (mmol/l), гликозилиран хемоглобин (HbA1c %) и статус на хипертензија. Прифатена е хибридна методолошка рамка која интегрира техники од применета статистика, визуелизација на податоци и примена на модели на машинско учење. Експлораторната анализа на податоците е проследена со примена на класификациски модели како што се логистичка регресија, класификатор на најблиски соседи (KNN) и дрвја на одлука. Резултатите откриваат статистички значајни корелации помеѓу дебелината и зголемени нивоа на холестерол и HbA1c, особено кај пациенти со дијабетес. Хипертензијата, исто така, покажува силна поврзаност со абнормални липидни профили. Компјутерското моделирање покажува дека БМИ, возраста и HbA1c се меѓу највлијателните предиктори за ризик од дијабетес.

Клучни зборови: Моделирање, Дијабетес мелитус, Напредна анализа на податоци.

1. Вовед

Дијабетес тип 2 настанува кога панкреасот не произведува доволно инсулин (хормон), или телото не го користи инсулинот правилно — или и двете. Ова е различно од дијабетес тип 1, кој настанува кога автоимуна реакција врз панкреасот резултира со целосен недостаток на производство на инсулин.

Дијабетес тип 2 најчесто се појавува кога телото има хипергликемија, односно постојано зголемено ниво на шеќер во крвта. Лицата со прекумерна маса (дебелина) се повеќе подложни, бидејќи се работи за складирани количини на калории кои телото не успеало да ги преработи и искористи. Дијабетесот е водечка причина за морбидитет и морталитет кај пациенти со кардиоваскуларни болести (КВБ). Лицата со дијабетес мелитус имаат 2 до 4 пати поголем ризик од КВБ во споредба со лица без дијабетес (Bruce et al., 2025). Дислипидемијата и хипертензијата се два главни фактори на ризик за КВБ што често коегзистираат со дијабетесот (Li et al., 2023). Истовремената појава на фактори на ризик за КВБ, како што се дислипидемија, хипергликемија, хипертензија и дебелина кај лица со дијабетес тип 2, уште повеќе ја зголемува нивната подложност на кардиоваскуларни болести (IDF, 2015 и Li X et al., 2023).

Фактите и бројките за дијабетесот ја прикажуваат растечката глобална тежина за поединците, семејствата и државите. Најновиот Атлас за дијабетес на IDF, 2025 (Bruce et al., 2025) известува дека 11,1% од возрасната популација живее со дијабетес, при што повеќе од 40% не се свесни дека ја имаат болеста. До 2050 година, се предвидува речиси 1 од 8 возрасни лица да живеат со дијабетес, при што повеќе од 90% од нив ќе имаат дијабетес тип 2. Главните фактори кои придонесуваат за порастот на дијабетес тип 2 се:

- Урбанизација
- Стареење на популацијата
- Намалено ниво на физичка активност
- Зголемена застапеност на прекумерна тежина и дебелина

Конечно, науката напредува и носи надеж во истражувањата за третман на лицата со дијабетес, во три насоки: технолошка, биолошка и фармаколошка. Технолошките решенија се насочени кон вештачки панкреас, биолошките кон трансплантација и регенерација на бета-клетки, а фармаколошките истражувања се најветувачки за значително намалување на товарот од дијабетес тип 2.

Менаџирањето на дијабетесот според ABC е особено важно:

- А (A1C тестот): просечен ниво на гликоза во последните 3 месеци;
- В (Blood pressure): контрола на крвен притисок;
- С (Cholesterol): управување со нивото на холестерол за превенција на срцеви заболувања.

За подобро предвидување на појавата на T2DM, истражувачите сè повеќе применуваат машинско учење (ML) и математичко моделирање. Овие пристапи откриваат сложени интеракции меѓу метаболичките индикатори и овозможуваат рана дијагноза и превенција.

2. Преглед на литературата

Дислипидемијата и хипертензијата се два главни ризик-фактори за КВБ што често коегзистираат со дијабетесот (Li et al., 2023). Истовремената појава на дислипидемија, хипергликемија, хипертензија и дебелина кај лица со T2DM дополнително ја зголемува нивната подложност на кардиоваскуларни болести (IDF, 2015; Li X et al., 2023).

Долготрајната консумација на исхрана со малку растителни влакна, внес на храна со висока калориска вредност и висок гликемиски товар се значајни фактори на ризик за

пред-дијабетес (Bansal, 2015). Жените со гестациски дијабетес или синдром на полицистични јајници имаат зголемен ризик за развој на пред-дијабетес (Tarvonen et al., 2021; Bell et al., 2020).

Прекумерната телесна маса и дебелината се силни предиктори за развој на пред-дијабетес, а постои и силна корелација со физичката неактивност, нездравата исхрана (Mohajan & Kumar, 2023), стрес и замор (Knights&Millaku, 2023). Дополнително, пациентите со дијабетес имаат поголема веројатност за присуство на висок крвен притисок или висок холестерол, што го зголемува ризикот од срцев удар или мозочен удар.

Истражувањата за примена на математичко моделирање и машинското учење и примена нутриционизмот и обезитост (Knights et al., 2023; Knights et al., 2024) покажуваат ветувачки резултати: од предикција и симулација на телесна маса и биохемиски параметри, до оптимизација на алгоритми за класификација на болести поврзани со дебелината (Knights&Damjanovska, 2024). Примената на регресиони модели во предикција (Knights & Prchkovska, 2024), како и определување на фактори кои влијаат на телесната маса (Knights et al., 2024; Markovik et al., 2022), претставува значаен придонес за подобро разбирање на комплексната поврзаност меѓу варијаблите потребни за предикција, односно клиничките и биохемиските параметри и ризикот од развој на дијабетес.

Имајќи ги во предвид на овие сознанија, ова истражување применува математички модели и машинско учење на реални податоци почитувајќи го етичкиот кодекс на доверливост од спроведено истражување во Клиничката болница „Мајка Тереза“ во Скопје. Користени се модели на логистичка регресија, дрвја на одлука и K-најблиски соседи, за процена на предиктивната моќ на варијабли како БМИ, холестерол, HbA1c и хипертензија.

3. Методи и материјали

3.1 Прибирање на податоци

Податоците беа собрани од клинички евиденции во период од девет месеци и вклучуваат анонимизирани податоци за повеќе од 200 пациенти кои доаѓале на прегледи три пати. Тоа значи дека сетот на податоци содржи 723 реда и 9 колони. Овие лица беа следени според индикатори поврзани со развој на дијабетес. Забележаните карактеристики се:

- Number – број на прегледи на пациент со зголемена телесна маса/дебелина, или обезитост/.
- HbA1c (%) – Гликозилиран хемоглобин што го претставува просечниот ниво на шеќер во крвта во последните три месеци; вредност $\geq 6.5\%$ е дијагностички маркер за дијабетес.
- Diastolic_pressure (mmHg) – Дијастолен крвен притисок.
- Lipid risk – Индикатор за проценка на метаболичкиот ризик поврзан со липиди, пресметан како:

$$\text{Lipid risk} = 0.4 \times \text{BMI} + 0.6 \times \text{SkinThickness} \quad (1)$$

повисоки вредности укажуваат на поголема веројатност за дислипидемија и кардиоваскуларен ризик.

- Body Mass Index (BMI) – Индекс на телесна маса, мерка за телесна маст врз основа на висина и тежина, пресметана како:

$$BMI = \frac{\text{Weight (kg)}}{(\text{Height (m)})^2} \quad (2)$$

BMI над 25 (kg/m²) укажува на прекумерна тежина; над 30 укажува на дебелина.

- Family_history – Педигре функција за дијабетес што ја квантифицира генетската предиспозиција врз основа на семејна историја. Во практика, се третира како нормализиран резултат меѓу 0 и 2.5.

$$\text{Праг на ризик: } \begin{cases} < 0.5 & \Rightarrow \text{Низок генетски ризик} \\ \geq 0.5 & \Rightarrow \text{Зголемен наследен ризик} \end{cases} \quad (3)$$

- Age – Возраст (години) на пациентот во моментот на евидентирање на податоците.
- Outcome (0 или 1) – Финален резултат на дијагноза на дијабетес, каде: 1 = Лицето има дијабетес, 0 = Лицето нема дијабетес.

3.2 Предобработка на податоци

Извршено е чистење на податоците за отстранување на нецелосни записи и корекција на неконзистентни формати. Континуираните променливи се стандардизирани (просек = 0, стандардна девијација = 1), а категоријалните променливи (на пр. дијагноза на дијабетес, пол) се кодирани преку one-hot encoding. Отстапувањата се откриени со IQR методот и третирани со винсоризација.

$$Z = \frac{x^i - \mu}{\sigma} \quad (4)$$

3.3 Графичка презентација на анализа на податоци

Користени се статистички прегледи, хистограми, boxplot-и и корелациски топлински мапи за визуелизација на распределбата и врските меѓу променливите. Пресметани се Пирсон и Спирман корелациски коефициенти за идентификација на асоцијации меѓу предикторите и исходот на дијабетесот.

3.4 Имплементација на модели

Применети се три супервизирани модели на класификација:

- Логистичка регресија: Базен модел што ја проценува веројатноста за дијабетес како функција од предикторските променливи со користење на сигмоидна функција.
- k-Најблиски соседи (KNN): Класифицира примероци врз основа на блискост до k-најблиски примери во тренинг сетот.
- Дрво на одлуки: Модел базиран на дрво што ги дели податоците врз основа на праг вредности на карактеристиките за да ја максимизира информациската добивка.

2.5 Метрики за евалуација

Перформансите на моделите беа евалуирани користејќи:

Точност (Accuracy), прецизност (Precision), поврат (Recall), F1-скор, и ROC криви (Receiver Operating Characteristic).

Сите модели беа имплементирани во **Python** користејќи ја библиотеката **scikit-learn**. Податоците беа поделени во однос 70:30 за обука и тестирање.

2.5 Метрики за евалуација

Перформансите на моделите беа евалуирани користејќи:

Точност (Accuracy), прецизност (Precision), поврат (Recall), F1-скор, и ROC криви (Receiver Operating Characteristic).

Сите модели се пресметувани во Python користејќи ја библиотеката **scikit-learn**. Податоците беа поделени во однос 70:30 за обука и тестирање.

4. Резултати и дискусија

4.1 Дескриптивна статистика и корелации

Табела 1 прикажува статистички преглед на сите променливи користени во множеството на податоци за предвидување на дијабетес. Вклучува: број на записи, просек, стандардна девијација, минимум, квартали (25%, 50%, 75%) и максимум за секоја карактеристика.

	Number	HbA1c	Diastolic _pressure	Lipid_ risk	Sex	BMI	Family_ history	Age	outc ome
count	200	200	200	200	200	200	200	200	200
mean	4,24	119,22	68,80	19,40	0,50	31,35	0,45	33,24	0,36
std	2,65	34,65	18,45	15,26	0,50	7,93	0,30	11,35	0,48
min	1,00	0,00	0,00	0,00	0,00	0,00	0,08	21,00	0,00
25%	2,00	96,00	62,00	0,00	0,00	27,20	0,25	24,00	0,00
50%	4,00	112,00	70,00	22,00	0,00	31,60	0,36	30,00	0,00
75%	6,00	143,50	78,00	32,00	1,00	35,45	0,59	40,00	1,00
max	15,00	197,00	122,00	51,00	1,00	55,00	2,33	69,00	1,00

Табела 1: Статистички преглед на користени променливи

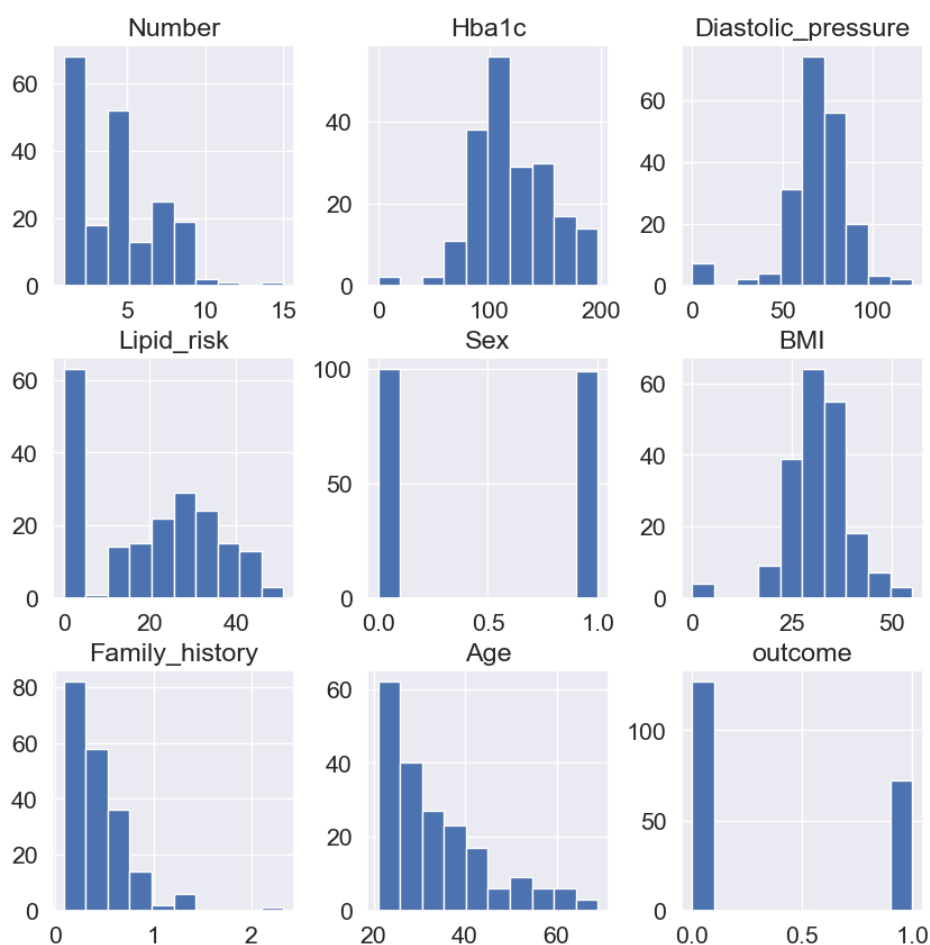
Бројот на прегледи (Number) има просек од 4,24, со медијана 4 и максимална вредност 15. Ова укажува дека поголемиот дел од пациентите имале релативно мал број медицински прегледи. HbA1c има средна вредност од 119 mmol/mol, со медијана 112 и распон од 0 до 197, што укажува на значителна варијабилност во гликемиската контрола и присуство на пациенти со многу високи вредности.

Кај дијастолниот притисок средната вредност е околу 69 mmHg, со медијана 70 и максимум 122, што одговара на вообичаени, но и на некои патолошки вредности. Липидниот ризик има средна вредност 19,4, медијана 22 и максимум 51, што укажува на поширок спектар на метаболички состојби.

За променливата Sex просекот е приближно 0,49, што потврдува дека примерокот е приближно рамномерно распределен помеѓу мажи и жени. Индексот на телесна маса (BMI) има просек од 31,35, со медијана 31,6, што покажува дека најголемиот дел од испитаниците се со прекумерна тежина или дебелина.

Кај Family_history просекот е 0,45, што укажува дека скоро половина од испитаниците имаат позитивна семејна анамнеза за дијабетес. Возраста (Age) се движи од 21 до 69 години, со медијана 30 и средна вредност 33, што ја претставува популацијата како релативно млада, но со присутни и постари пациенти.

Сликата 1 ги прикажува хистограмите за девет клинички варијабли кои се земени како потенцијални предиктори на дијабетес тип 2: број на прегледи, HbA1c, дијастолен крвен притисок, липиден ризик, ниво на инсулин, индекс на телесна маса (BMI), семејна анамнеза, возраст и исход (0 = нема дијабетес, 1 = дијабетес).

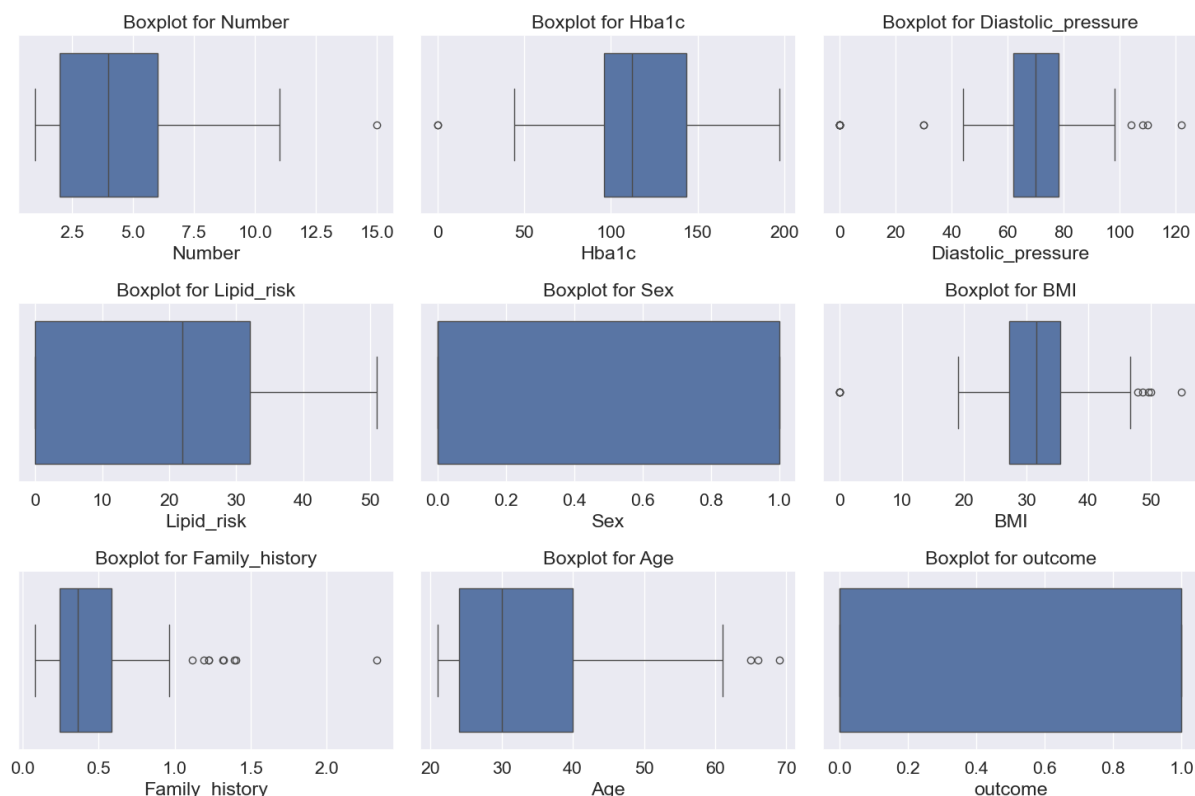


Слика бр. 1: Хистограм на распределба на карактеристиките (варијабли) кај испитаниците

Забележливо е дека повеќето испитаници имаат мал број на посети и дека вредностите на HbA1c се концентрирани околу 100 mmol/mol, што укажува на хипергликемија. Дијастолниот притисок најчесто се движи меѓу 60–80 mmHg. Дадена е и променливата Sex, која е бинарна (0 = женски, 1 = машки пол), се гледа дека е рамномерна распределба, што укажува дека во примерокот има скоро подеднакво

застапени мажи и жени. Дистрибуцијата на BMI е приближно нормална со доминантни вредности околу 25–35, што укажува на прекумерна тежина и дебелина кај голем дел од популацијата. Кај променливата возраст се забележува дека мнозинството испитаници се на возраст од 20 до 40 години, додека исходот покажува дека иако поголем дел од испитаниците немаат дијабетес, значителен број веќе се со дијагностицирана состојба. Овие резултати ја потврдуваат важноста на метаболичките и демографските фактори како предиктори за дијабетес тип 2.

Boxplot графичите овозможуваат визуелизација на медијаната, интерквartilниот опсег (IQR), како и идентификација на потенцијални екстремни вредности (outliers).



Слика бр. 2: Boxplot дијаграми за сите вклучени променливи во анализата

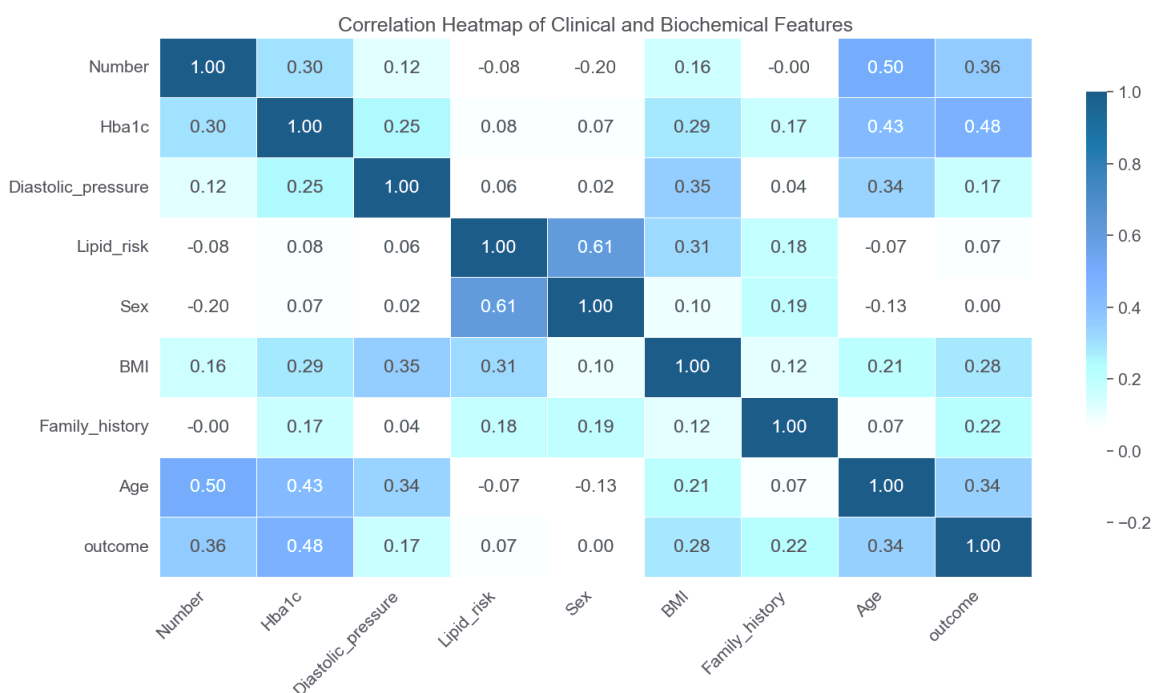
Кај променливата Number (број на прегледи) повеќето вредности се концентрирани до пет, додека поединечни пациенти имаат многу високи вредности над десет, што претставува outliers. Кај HbA1c медијаната се наоѓа околу 100 mmol/mol, што укажува на присутна хипергликемија кај мнозинството испитаници, но со поединечни случаи со значително пониски и повисоки вредности. Кај Diastolic_pressure најчести вредности се во интервалот 60–80 mmHg, а присутни се и неколку екстреми на ниски и високи вредности.

Lipid_risk покажува поширок интерквartilен опсег, со медијана околу 20–25 и екстремни вредности блиску до 50. Променливата Sex е бинарна (0 и 1), при што нема отстапки, што значи дека примерокот е добро избалансиран. Кај BMI медијаната е околу 30, што укажува на прекумерна тежина или дебелина како доминантна состојба, а видливи се неколку екстреми над 45–50.

Family_history има медијана околу 0.5, што укажува дека дел од пациентите имаат позитивна семејна историја за дијабетес, со присутни поединечни екстреми кај вредности повисоки од 1. Кај Age најголемиот број испитаници се меѓу 25 и 40 години,

а екстремни вредности се јавуваат кај лица над 60. Конечно, кај outcome (0 = нема дијабетес, 1 = дијабетес) се гледа јасна бинарна распределба без значајни отстапки.

На слика 3 е претставена топлинака мапа на коефициентите на корелација помеѓу променливите. Забележливо е дека постои умерена позитивна корелација помеѓу HbA1c и исходот (outcome, $r=0,48$), што потврдува дека зголемените вредности на HbA1c се силен индикатор за присуство на дијабетес тип 2. Слично, бројот на прегледи (Number) и возраста (Age) имаат позитивна поврзаност со исходот ($r=0,36$ и $r=0,34$), што укажува дека постарите пациенти и оние со повеќе медицински посети имаат поголема веројатност за дијабетес.



Слика бр. 3: Топлинска мапа на коефициент на корелација помеѓу променливите

Дијастолниот крвен притисок има умерена корелација со BMI ($r=0,35$), што е очекувано бидејќи зголемената телесна маса често води до хипертензија. Исто така, постои умерена поврзаност меѓу Lipid_risk и Sex ($r=0,61$), што може да укажува на разлики во липидниот профил меѓу половите. Возраста е позитивно поврзана со повеќе променливи, вклучително и HbA1c ($r=0,43$) и дијастолен притисок ($r=0,34$), што ја нагласува улогата на стареењето како ризик-фактор за метаболички и кардиоваскуларни нарушувања.

Од друга страна, Family_history покажува само слаба линеарна поврзаност со исходот ($r=0,22$), што значи не значи дека иако генетската предиспозиција има слаба улога, напротив треба да се испитаат не линеарни односи.

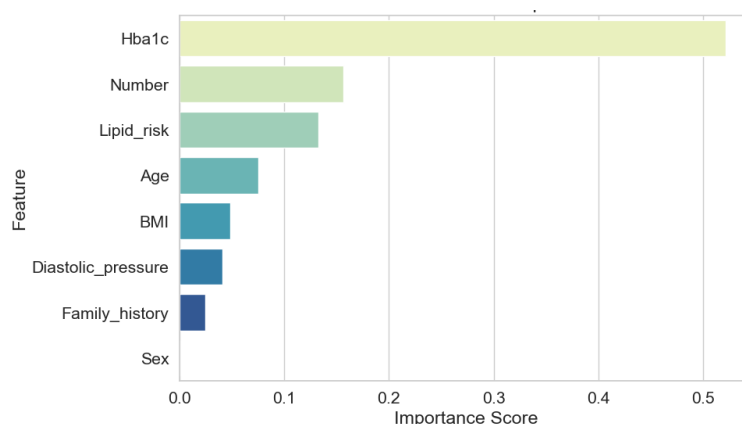
Вкупно, топлинската мапа открива дека HbA1c, бројот на прегледи, возраста и BMI се најзначајни фактори кои се директно или индиректно поврзани со исходот на дијабетес тип 2, и затоа тие треба да бидат вклучени во понатамошното предиктивно моделирање.

4.2 Фактори на вијание

Поради сложените метаболички процеси линерните корелации не се доволни за да се прикажат вистинските релации на факторите кои вијаат на меѓусебна поврзаност на паромениливите. За таа цел спроведени се неколку модели на предикција користејќи машинско учење: логистичка регресија, decision tree (дрво на одлучување) и K-Nearest Neighbors (KNN) (K-најблиски соседи).

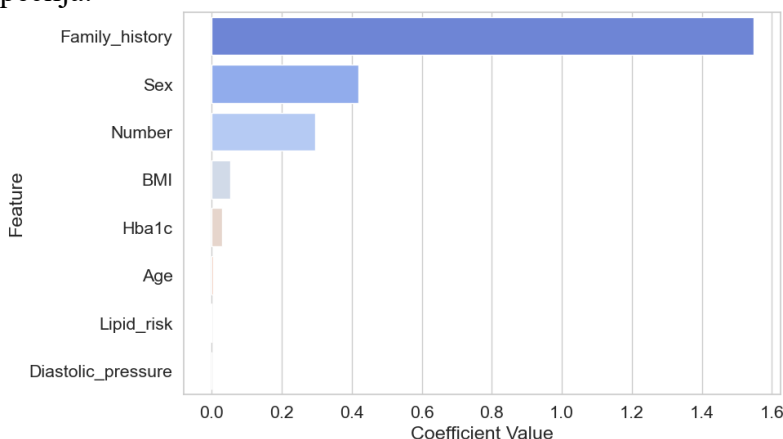
Според овие модели добиени се следните фактори на вијание во предикција и поврзаност на дијабетес и обезноста.

Слика 4, ја прикажува важноста на клиничките и биохемиските параметри во моделот -дрво на одлука за предвидување на дијабетес тип 2. Највисока важност има променливата HbA1c (над 0,5), што укажува дека контролата на гликемија е најсилниот индикатор за присуство на болеста. Следуваат бројот на медицински прегледи (Number) и липидниот ризик (Lipid_risk), кои имаат умерено висока предиктивна моќ. Возраста (Age) и индексот на телесна маса (BMI) имаат средна важност, што ја потврдува нивната улога како ризик-фактори. Дијастолниот притисок (Diastolic_pressure), семејната историја (Family_history) и полот (Sex) имаат најниска важност, што сугерира дека нивното влијание врз предвидувањето во овој модел е ограничено.



Слика бр. 4: Фактори на вијание во моделот - дрво на одлука

На сликата 5 е прикажана важноста на факторите (променливите) во моделот на логистичка регресија.

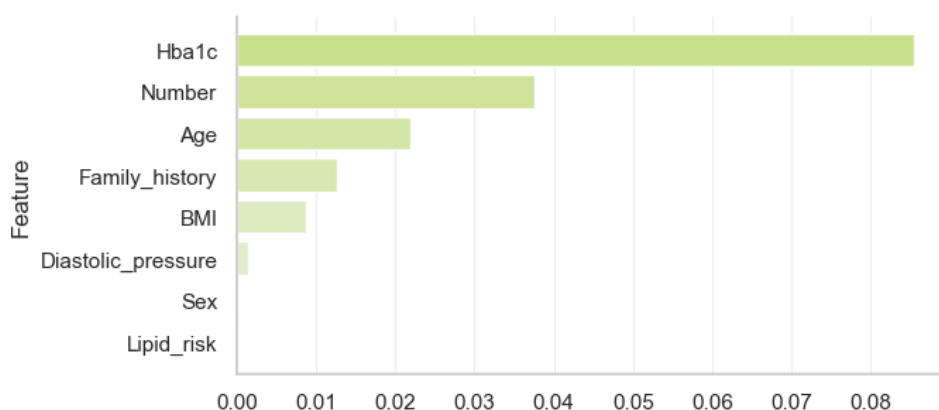


Слика бр. 5: Фактори на вијание во моделот-логичка регресија

Најсилен позитивен предиктор за дијабетес е *Family_history*, што укажува дека семејната анамнеза има клучно значење за ризикот. Следат *Sex* и *Number* (број на медицински прегледи), кои исто така имаат значаен позитивен придонес. Индексот на телесна маса (*BMI*) и *HbA1c* имаат помал, но сепак позитивен ефект, додека возраста (*Age*), липидниот ризик (*Lipid_risk*) и дијастолниот притисок (*Diastolic_pressure*) имаат минимално влијание во рамките на овој модел.

Овие резултати покажуваат дека логистичката регресија повеќе ја нагласува улогата на наследните и демографските фактори (семејна историја, пол), за разлика од моделот на дрво на одлука кој ја истакнуваше биохемијата (*HbA1c*).

На слика 6, прикажани се факторите на вијание добиени со моделот-К-најблиски соседи (KNN)



Слика бр. 6: Фактори на вијание во моделот-К-најблиски соседи (KNN)

Овие резултати (слика 6) покажуваат дека метаболичкиот маркер *HbA1c* е доминантен индикатор за предвидување на дијабетес, додека останатите варијабли имаат дополнителна, но многу помала предиктивна моќ. KNN моделот, кој се базира на сличност меѓу пациенти, најмногу се потпира на објективните лабораториски мерења, а многу помалку на демографските или клиничките ризик фактори.

4.3 Споредба на моделите

За да се оцени перформансот на различните машински модели за предвидување на дијабетес, беа обучени и евалуирани три модели:

- Логистичка Регресија
- Decision Tree (пред и по подесување на хиперпараметри)
- K-Nearest Neighbors (KNN) (пред и по подесување)

Споредбата на моделите се базира на следните метрики: точност (accuracy), прецизност, поврат (recall) и F1-резултат.

Резултатите (Табела 2) покажуваат значајни разлики во перформансите помеѓу трите модели.

Дрво на одлуки: Овој модел постигна најдобри резултати во споредба со останатите. Точноста изнесува 96.1%, а истата вредност е добиена и за AUC. Исто така, прецизноста, повратот и F1-резултатот се високи (0.942), што укажува на добро

балансирана способност на моделот и за точна класификација и за откривање на позитивните случаи (пациенти со дијабетес). Ова значи дека моделот е и прецизен и чувствителен, без да жртвува една метрика на сметка на друга.

Модел	Точност	Прецизност	Поврат	F1	AUC
Дрво на одлуки	0.961	0.942	0.942	0.942	0.961
Логистичка Регресија	0.786	0.880	0.423	0.571	0.786
KNN	0.779	0.714	0.577	0.638	0.779

Табела 2. Преглед на перформансите на моделите

Логистичка регресија: Иако е едноставен и често користен модел за бинарна класификација, резултатите овде се послаби. Точноста и AUC изнесуваат 78.6%, а особено низок е повратот (0.423). Тоа значи дека моделот има тенденција да пропушта поголем дел од вистинските позитивни случаи (false negatives). Од друга страна, прецизноста е висока (0.880), што значи дека кога класифицира некого како позитивен, тоа во голем дел е точно. Сепак, слабата чувствителност ја намалува неговата практична примена за медицинска дијагностика, каде е важно да не се пропуштат пациенти.

K-Nearest Neighbors (KNN): покажува слични резултати со логистичката регресија, со точност и AUC од 77.9%. Прецизноста е умерена (0.714), но повратот (0.577) е нешто подобар од оној на логистичката регресија. F1-резултатот (0.638) укажува на подобар баланс меѓу прецизност и поврат во споредба со логистичката регресија. Овој модел е нешто поуспешен во откривање на позитивните случаи, но сè уште значително послаб од дрвото на одлуки.

5. Заклучок

Резултатите од ова истражување покажуваат дека примената на математички модели и техники на машинско учење овозможува ефикасна предикција на дијабетес тип 2 врз основа на клинички и биохемиски параметри. Најзначајни предиктори за развој на болеста се гликозилираниот хемоглобин (HbA1c), индексот на телесна маса (BMI), возраста и липидниот ризик, што е во согласност со досегашните клинички сознанија.

Во споредбата на моделите, дрвото на одлуки покажа најдобра перформанса, со највисока точност, прецизност, поврат и AUC, што го прави најсоодветен за идентификација на пациенти со зголемен ризик од дијабетес тип 2. Логистичката регресија и KNN, иако постигнаа солидни резултати, имаа пониска чувствителност и рамнотежа меѓу метриците, што ја ограничува нивната применливост.

Користена литература

Ahmad, H., Yousaf, S., Samina, M., Naqvi, R. A., Khan, A., & Rizwan, S. (2025). Non-high-density lipoprotein cholesterol: A marker of cardiovascular disease in type-2 diabetes mellitus. *Pakistan Journal of Pathology*, 36(1). <https://doi.org/10.55629/pakjpathol.v36i1.843>

Bansal, N. (2015). Prediabetes diagnosis and treatment: A review. *World Journal of Diabetes*, 6(2), 296–303.

Bell, K., et al. (2020). A position statement on screening and management of prediabetes in adults in primary care in Australia. *Diabetes Research and Clinical Practice*, 164, 108188.

International Diabetes Federation (IDF). (2015). *IDF Diabetes Atlas* (7th ed.). Brussels: International Diabetes Federation. https://diabetesatlas.org/idfawp/resource-files/2012/07/IDF_diabetes_atlas_eventh_edition_en.pdf

International Diabetes Federation (IDF). (2024). What we do – education. <https://idf.org/what-we-do/education/>

International Diabetes Federation (IDF). (2025). *IDF Diabetes Atlas* (11th ed.). Brussels: International Diabetes Federation.

Knights, V., Kolak, M., Markovikj, G., & Gajdoš Kljusurić, J. (2023). Modeling and optimization with artificial intelligence in nutrition. *Applied Sciences*, 13(13), 7835. <https://doi.org/10.3390/app13137835>

Knights, V., Blazevska, T., Markovic, G., & Kljusurić, J. (2024a). Mathematical analysis of statistical design of experiment and machine learning methods in identifying factors influencing obesity. *Medical Research Archives*, 12(9). <https://doi.org/10.18103/mra.v12i9.5790>

Knights, V., & Gavriloska, E. (2024). Machine learning techniques for modelling and predicting the influence of kefir in a low-protein diet on kidney function. *Medical Research Archives*, 12(7). <https://doi.org/10.18103/mra.v12i7.5631>

Knights, V., & Prchkovska, M. (2024). From equations to predictions: Understanding the mathematics and machine learning of multiple linear regression. *Journal of Mathematical & Computer Applications*, SRC/JMCA-168. [https://doi.org/10.47363/JMCA/2024\(3\)137](https://doi.org/10.47363/JMCA/2024(3)137)

Knights, V., & Millaku, J. (2023). Three-factor experimental design as a tool in applied statistics. *International Journal of Statistics and Applied Mathematics*, 8(1), 46–49. <https://doi.org/10.22271/math.2023.v8.i1a.929>

Li, X., Shi, Y., Wei, D., Ni, W., Zhu, N., & Yan, X. (2023). Impact of a high dietary fiber cereal meal intervention on body weight, adipose distribution, and cardiovascular risk among individuals with type 2 diabetes. *Frontiers in Endocrinology*, 14, 1283626. <https://doi.org/10.3389/fendo.2023.1283626>

Markovikj, G., & Knights, V. (2022). Model of optimization of the sustainable diet indicators. *Journal of Hygienic Engineering and Design*, 39, 169–175.

Markovikj, G., Knights, V., & Gajdoš Kljusurić, J. (2023). Ketogenic diet applied in weight reduction of overweight and obese individuals with progress prediction by use of the modified Wishnofsky equation. *Nutrients*, 15(4), 927. <https://doi.org/10.3390/nu15040927>

Mohajan, D., & Mohajan, H. K. (2023). Prevention and management strategies of pre-diabetes. *Fundamental Management Studies*, 10(4). <https://doi.org/10.56397/FMS.2023.10.04>

National Institute of Diabetes and Digestive and Kidney Diseases (NIDDK). (2024). Preventing heart disease and stroke. <https://www.niddk.nih.gov/health-information/diabetes/overview/preventing-problems/heart-disease-stroke>

Shomali, M. (2012). Diabetes treatment in 2025: Can scientific advances keep pace with prevalence? *Therapeutic Advances in Endocrinology and Metabolism*, 3(4), 163–173. <https://doi.org/10.1177/2042018812465639>

Tarvonen, M., et al. (2021). Intrapartal cardiotocographic patterns and hypoxia-related perinatal outcomes in pregnancies complicated by gestational diabetes mellitus. *Acta Diabetologica*, 58(11), 1563–1573.