

**ՏԵՂԵԿԱՏՎԱԿԱՆ ՏԵԽՆՈԼՈԳԻԱՆԵՐ, ԷԼԵԿՏՐՈՆԻԿԱ,  
ՌԱԴԻՈՏԵԽՆԻԿԱ**

**ԻՆՖՈՐՄԱՏԻԿԱ ԵՎ ՀԱՇՎՈՂԱԿԱՆ ՏԵԽՆԻԿԱ**

ՀՏԴ 681.3

**Լ.Կ. ԱՆԴՐԵԱՍՅԱՆ**

**ՏՎՅԱԼՆԵՐԻ ԴԱՍԱԿԱՐԳՄԱՆ ՑՈՒՑԱՆԻՇՆԵՐԸ ԿԱՆԽԱՐԳԵԼԻՉ  
ԲԺՇԿՈՒԹՅԱՆ ՈՐՈՇՈՒՄՆԵՐԻ ԿԱՅԱՑՄԱՆՆ ԱՋԱԿՑՈՂ ՀԱՄԱԿԱՐԳՈՒՄ**

Ներկայացվել է կանխարգելիչ բժշկության որոշումների կայացմանն աջակցող համակարգում 2-րդ տիպի շաքարային դիաբետ (ՏՇԴ) հիվանդության ախտորոշման համար տվյալների դասակարգման մոդելների ճշտության ցուցանիշների համեմատումը:

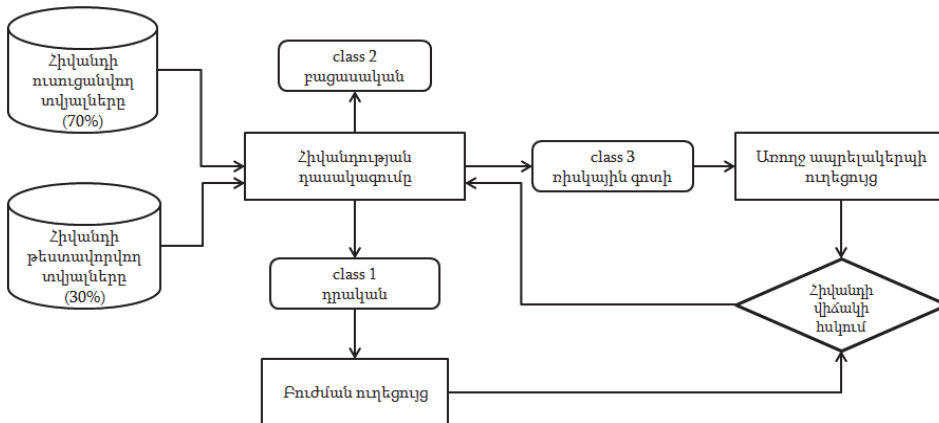
**Առանցքային բառեր:** վերահսկելի մեքենայական ուսուցում, 2-րդ տիպի շաքարային դիաբետ, տվյալների դասակարգման ցուցանիշներ:

**Ներածություն.** Կանխարգելիչ բժշկության որոշումների կայացմանն աջակցող համակարգում (ԿԲՈԿԱՀ) վերահսկելի մեքենայական ուսուցման մեթոդներով դասակարգումը հիվանդների վերաբերյալ հավաքված անձնական, բժշկական հետազոտությունների և բուժման արդյունքների հավաքածուի տարաբնույթ ու բարդ տվյալների մշակման գործառույթ է, որի արդյունքում ստացված մոդելը, կիրառվելով ըստ բժշկական տվյալների, կարող է կատարել հիվանդության ախտորոշում, առաջարկել բուժման սխեմաներ (նկ.1) [1]: Համակարգում որպես տվյալների հավաքածու օգտագործվել է NHANES (National Health and Nutrition Examination Survey) 2013-2014թթ.-ին կատարված հարցումների հավաքածուն, որը ներառում է հարցման մասնակիցների ֆիզիկական և լաբորատոր հետազոտությունները, ինչպես նաև սննդի կարգավիճակի վերաբերյալ տվյալները [2]: Մուտքային հատկանիշների համար ընտրվել են.

- մասնակիցների անձնական տվյալներից՝ սեռը (*Gender - 1-male` արական, 2-female` իգական սեռ*), տարիքը (*Age*), քաշի և հասակի քառակուսու հարաբերությունը (*Body Mass Index - BMI*),

- լաբորատոր հետազոտությունների արդյունքներից՝ ընդհանուր խոլեստերինը (*Total Cholesterol - TCI*), գլիկոգլիացված հեմոգլոբինը (*GlycoHemoglobin - H1Ac*), արյան ճնշման սիստոլիկ (*Systolic Blood Pressure - SBP*) և դիաստոլիկ (*Diastolic Blood Pressur - DBP*) ցուցանիշները:

Համեմատվել են վերահսկելի մեքենայական ուսուցման **Linear Regression (LR)**, **K-Nearest Neighbors (KNN)**, **Multilayer Perceptron (MLP)** դասակարգման և **Adaboost**, **Bagging Decision Tree (BDT)**, **Random Forests (RFs)**, **Extra Trees (ETs)** անսամբլային մոդելների արդյունավետության հետևյալ չափանիշները՝ **Accuracy**, **Sensitivity**, **Specificity** [2, 3]:



Նկ. 1. ԿԲՈԿԱՀ-ում տվյալների դասակարգման սխեման

Դասակարգման արդյունքում տվյալները բաշխվել են ըստ S2CԴ հիվանդության առկայության (դրական), բացառման (բացասական) և ռիսկային գոտում գտնվելու երեք դասերի՝ **class 1**, **class 2**, **class 3** (նկ.1), որի դեպքում ունենում ենք եռաչափ սխալների մատրիքս (**confusion matrix**)՝ անկյունագծերում տեղակայված ճշգրիտ դասակարգված նմուշներով (նկ.2):

Կանխատեսված

		Class 1	Class 2	Class 3
Փաստացի	Class 1	Ճշգրիտ դասակարգված		
	Class 2		Ճշգրիտ դասակարգված	
	Class 3			Ճշգրիտ դասակարգված

Նկ. 2. Սխալների մատրիքսը 3 դասերի դեպքում

Նշանակենք մատրիքսը  $X = \{x_{ij}\}$   $i, j = 1, 2, 3$  համար, որտեղ  $x_{ij}$ -երը  $i$ -րդ դասին պատկանող նմուշներն են, որոնք կանխատեսվել են որպես  $j$ -րդ դաս: Ներկայացնենք.

$$\text{True Positives} - tp_i = x_{ii}, \quad (1)$$

$$\text{False Positives} - fp_i = \sum_{j=1,3} x_{ji} - tp_i, \quad (2)$$

$$\text{False Negatives} - fn_i = \sum_{j=1,3} x_{ij} - tp_i, \quad (3)$$

$$\text{True Negatives} - tn_i = \sum_{j=1,3} \sum_{k=j,3} x_{jk} - tp_i - fp_i - fn_i, \quad (4)$$

որտեղ  $x_{ii}$ -ը ճշգրիտ կանխատեսված նմուշներն են [1]: Այդ դեպքում **Accuracy**-ին, **Sensitivity** և **Specificity**-ին հաշվարկվում են հետևյալ բանաձևերով.

$$\text{Accuracy} = \sum_{i=1,3} (tp_i + tn_i) / (tp_i + tn_i + fp_i + fn_i), \quad (5)$$

$$\text{Sensitivity}_i = tp_i / (tp_i + fn_i), \quad (6)$$

$$\text{Specificity}_i = tn_i / (tn_i + fp_i): \quad (7)$$

**Խնդրի դրվածքը.** Ներկայացնել S2ՇԴ-ի հիվանդության ախտորոշման համար բժշկական տվյալների դասակարգման համար օգտագործված **LR, KNN, MLP և Adaboost, BDT, RFs, ETs** անսամբլային վերահսկելի մեքենայական ուսուցման մոդելների ճշտության ցուցանիշների համեմատումը:

**Ուսուցանվող հավաքածուի նախապատրաստումը.** NHANES (2013-2014թթ.)-ի տվյալների հավաքածուի հիման վրա, որը ներառում է 9813 հետազոտված մարդկանց գրառումներ, ուսուցանվող հավաքածուի նախապատրաստման համար կատարվել են հետևյալ գործառնությունները [4].

- բացակայող (**missing**) տվյալները փոխարինվել են միջին արժեքով,
- շեղված (**outlier**) փոփոխականներ ունեցող տվյալներն անտեսվել են,
- կրկնվող (**duplicate**) տվյալները միավորվել են,
- անիմաստ (**noise data**) տվյալները զտվել են,
- հեռացվել են 40-ից փոքր տարիք ունեցողների տվյալները:

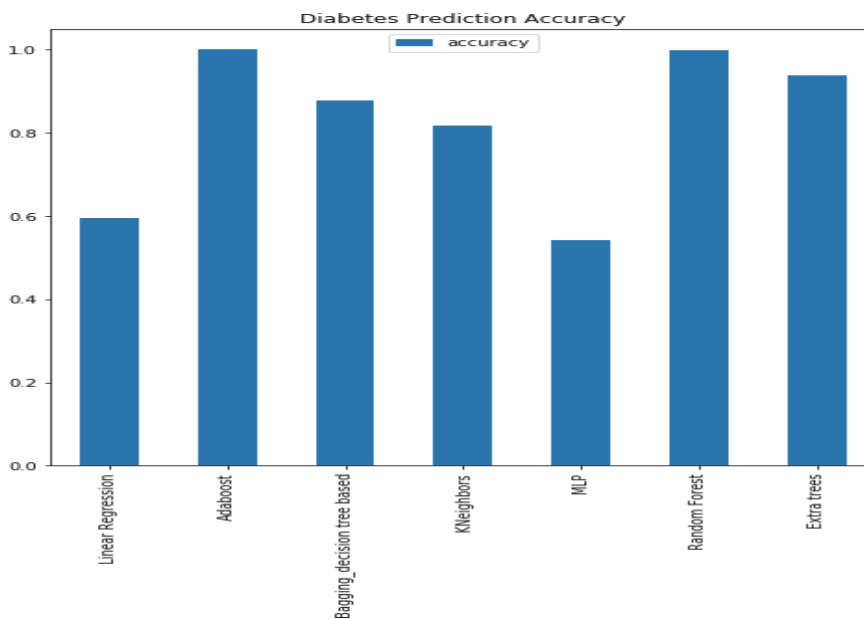
Զտված հավաքածուի վերջնական տարբերակում 1943 գրառումները կանանց տվյալներն են, իսկ 1737-ը՝ տղամարդկանց:

**Դասակարգման մոդելների ճշտության ցուցանիշների համեմատումը.** Զտված հավաքածուի տվյալների դասակարգման համար վերոհիշյալ մոդելների աշխատանքի արդյունքում ստացված ճշտության ցուցանիշները ներկայացված են աղ. 1-ում և գրաֆիկորեն պատկերված են նկ. 3-ում:

Մոդելների ճշտությունները

	Մոդելը	Ճշտությունը (Accuracy)
0	Linear Regression	0.594876
1	Adaboost	1.000000
2	Bagging_decision tree based	0.966176
3	KNN	0.891176
4	MLP	0.541176
5	Random Forest	0.998529
6	Extra Trees	0.936765

Մոդելների աշխատանքի արդյունքների արդյունավետությունը գնահատելու համար կիրարկվել է  $k$  խաչաձև վավերացման մեթոդը (***k-fold cross validation***), ըստ որի հավաքածուն բաժանվում է  $k$  ենթաբազմությունների,  $k-1$  մասն օգտագործվում է ուսուցման, իսկ  $1/k$  մասը՝ թեստի համար, յուրաքանչյուր իտերացիայի ժամանակ փոխվում է թեստի համար նախատեսված տվյալների ենթաբազմությունը:



Նկ. 3. Մոդելների ճշտությունների գրաֆիկական պատկերը

Ներկայացված մոդելների համար հաշվարկվել են սխալների մատրիքները, ճշտությունը (**Accuracy**), **Sensitivity**-ն, այսինքն՝ հավանականությունը, որ

արդյունքը կլինի դրական՝ հիվանդության առկայության դեպքում (**True Positive Rate**) և **Specificity**-ն, այսինքն՝ հավանականությունը, որ արդյունքը կլինի բացասական՝ հիվանդության բացակայության դեպքում (**True Negative Rate**) [5, 6]: Մոդելներով տվյալների դասակարգման ցուցանիշների արդյունքների համեմատման համար, ըստ սխալների մատրիքի, հաշվարկվել են միջին քառակուսային շեղումները (աղ. 2):

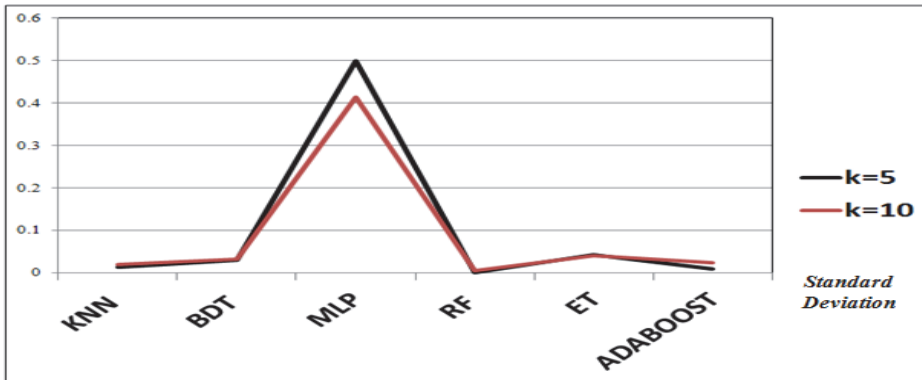
Տարբեր պարամետրերով  $k=5$  և  $k=10$  խաչաձև վավերացման մեթոդը կիրարկելուց հետո ըստ մոդելների միջին քառակուսային շեղումների պատկերված դիագրամի երևում է, որ ամենալավ արդյունքը գրանցվել է փոքր շեղում ունեցող անսամբլային մեթոդների (**ensemble learning – bagging and boosting**) կիրառման դեպքում, դրանք են՝ **BDT, Adaboost** և **RFs** հենքով մոդելները (նկ. 4):

Աղյուսակ 2

Մոդելների սխալների մատրիքը և ստանդարտ շեղումը

Confusion matrix of the KNN classifier	Confusion matrix of the BDT classifier	Confusion matrix of the MLP classifier
[[344 24 0] [ 11 262 0] [ 5 34 0]] Accuracy : 0.8912 Sensitivity : 0.9348 Specificity : 0.9597 Standart Deviation : 0.0125	[[364 4 0] [ 19 254 0] [ 0 0 39]] Accuracy : 0.9088 Sensitivity : 0.9891 Specificity : 0.9304 Standart Deviation : 0.0294	[[368 0 0] [ 0 273 0] [ 0 39 0]] Accuracy : 0.5412 Sensitivity : 1.0 Specificity : 0.0 Standart Deviation : 0.500
Confusion matrix of the RFs classifier	Confusion matrix of the ETs classifier	Confusion matrix of the Adaboost classifier
[[368 0 0] [ 1 272 0] [ 0 0 39]] Accuracy : 0.9412 Sensitivity : 1.0 Specificity : 0.9963 Standart Deviation : 0.0018	[[360 8 0] [ 29 244 0] [ 1 5 33]] Accuracy : 0.8882 Sensitivity : 0.9783 Specificity : 0.8938 Standart Deviation : 0.0422	[[368 0 0] [ 0 273 0] [ 0 0 39]] Accuracy : 0.943 Sensitivity : 1.0 Specificity : 1.0 Standart Deviation: 0.000

Աղյուսակում ներկայացված մոդելներով ըստ **Class 1, Class 2, Class 3** դասերի ճշգրիտ դասակարգված նմուշների քանակն է. **KNN** - 344, 262, 0; **BDT** - 364, 254, 39; **MLP** -368, 273, 0; **RFs** -368, 272, 39; **Ets** - 360, 244, 33; **Adaboost** - 368, 273, 39;



Նկ. 4. Մոդելների միջին քառակուսային շեղումների դիագրամը  $k=5$  և  $k=10$  խաչաձև վավերացման մեթոդի կիրառումից հետո

**Եզրակացություն.** Հետազոտության արդյունքներից կարելի է եզրակացնել, որ անսամբլային ալգորիթմները՝ որպես հիբրիդային ենթահամակարգեր, ապահովում են կանխատեսման առավել բարձր ճշգրտությունն ու ցածր սխալանքը: Նախատեսվում է այդ մոդելները կիրարկել NHANES 2011-2018թթ.-ին կատարված հարցումների մեծաքանակ տվյալների հավաքածուում, և կատարել ճշտության ցուցանիշների համեմատում:

#### ԳՐԱԿԱՆՈՒԹՅԱՆ ՑԱՆԿ

1. **Андреасян Л.К.** Классификация данных в системе принятия решений для профилактики сахарного диабета на основе методов машинного обучения //Вестник НПУА.-2018.- № 1.- С.36-44.
2. <https://www.cdc.gov/nchs/nhanes/index.htm>
3. **Андреасян Л.К.** Обзор моделей и подходов машинного обучения для профилактики и диагностики болезни сахарного диабета 2-го типа. ISSN 1829-0043 //ВЕСТНИК ИНЖЕНЕРНОЙ АКАДЕМИИ (ВИАА). – 2018.- Т.15, №% 1.- С.151-154.
4. **Liu H., Yu L.** Towards Integrating Feature Selection Algorithms for Classification and Clustering. //IEEE Transactions on Knowledge and Data Engineering. – 2005.- 17(3).- p. 1-12.
5. **Cleophas, Ton J., Zwinderman, Aeilko H.** Machine Learning in Medicine. - A Complete Overview, 2015.- 224p.
6. **Witten I.H., Eibe F., Hall M.A.** Data Mining: Practical Machine Learning Tools and Techniques. – Morgan Kaufmann, San Francisco, 2011.- 558 p.

Ջ.Կ. ԱՆԴՐԵԱՏՅԱՆ

**ПОКАЗАТЕЛИ КЛАССИФИКАЦИИ ДАННЫХ В СИСТЕМЕ  
ПОДДЕРЖКИ ПРИНЯТИЯ ПРЕВЕНТИВНЫХ МЕДИЦИНСКИХ  
РЕШЕНИЙ**

Представлено сравнение точности моделей классификации данных для профилактики болезни сахарного диабета 2-го типа (СД2Т) в системе поддержки принятия превентивных медицинских решений.

**Ключевые слова:** система поддержки принятия превентивных медицинских решений, контролируемое машинное обучение, сахарный диабет 2-го типа, показатели классификации данных.

L.K. ANDREASYAN

**DATA CLASSIFICATION METRICS IN THE PREVENTIVE MEDICAL  
DECISION MAKING SUPPORT SYSTEM**

The comparison of the accuracy metrics classification models of data prophylaxis for Type 2 Diabetes Mellitus (T2DM) in the preventive medical decision making support system is introduced.

**Keywords:** preventive medical decisions making support system, supervised machine learning, Type 2 Diabetes Mellitus, data classification metrics.

ՀՏԴ 681.3

Ա.Կ. ՄԱՅԻԼՅԱՆ, Ա.Ժ. ՄՈՄՋՅԱՆ

**JAVA ՄԻՋԱՎԱՅՐՈՒՄ WEB ԾՐԱԳՐԱՎՈՐՄԱՆ ՏԵԽՆՈԼՈԳԻԱՆԵՐԻ  
ՀԵՏԱԶՈՏՈՒՄԸ ԵՎ ՀԱՄԵՄԱՏԱԿԱՆ ՎԵՐԼՈՒԾՈՒԹՅՈՒՆԸ**

Հետազոտվել են Java միջավայրում web ծրագրավորման տեխնոլոգիաները, և կատարվել է դրանց համեմատական վերլուծությունը՝ ըստ հետևյալ պարամետրերի. արագ նախատիպային հավելված, օգտագործման պարզություն, փաստաթղթեր և հետադարձ կապ, թողունակություն և անսահմանափակություն:

**Առանցքային բաներ.** նախատիպ, նախագիծ, գործիք, շաբլոն, ինկապսուլյացիա, ինտերֆեյս, հղում, թողունակություն, դաս, կոնֆիգուրացիա:

Ծրագրավորողներից շատերը նախագիծ սկսելուց կանգնում են մի կարևորագույն խնդրի՝ ծրագրավորման լեզվի և framework-ի ընտրության առջև, որն անելու համար պետք է իմանալ, թե որ պարամետրերը համեմատել և որոշել՝ որն է առավելագույնս համապատասխանում կատարվելիք նախագծին: