
ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ

УДК 004.8

<https://doi.org/10.47533/2020.1606-146X.58>

*А. А. *МЫҢЖАСАР, А. Б. КУЛЬДЖАБЕКОВ,
Б. С. ДАРИБАЕВ, А. Н. ТЕМІРБЕКОВ*

Әл-Фараби атындағы Қазақ ұлттық университеті, Алматы қаласы, Қазақстан
myngzhassar_akniyet@mail.ru
alibek.kuljabekov@gmail.com
beimbet.daribayev@gmail.com

МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІН ПАЙДАЛАНУ АРҚЫЛЫ АДАМДАРДЫҢ MYERS-BRIGGS TYPE INDEX (MBTI) ТИПІН АНЫҚТАУ

Мақала компьютерлік лингвистика саласындағы машиналық оқыту мәселелеріне, атап айтқанда адамдардың психологиялық типтерін әлеуметтік желідегі мәтіндік жазбалары негізінде анықтауға негізделген. Бұл мақаланың мақсаты – адамдарды әлеуметтік желідегі жазбаларының мәтін үлгілері негізінде Myers-Briggs Type Index (MBTI) типін анықтайтын қазақ тіліне арналған классификатор құру үшін машиналық оқытудың Naive bayes және Extreme Gradient Boosting(XGBoost) әдістерін зерттеу болып табылады. Машиналық оқыту әдістерін пайдаланудағы зерттеу тәжірибелерінің барысы мен зерттеу нәтижелері келтіріліп, алынған нәтижелер салыстырылды.

***Түйін сөздер:** Naive Bayes, XGBoost, TF-IDF, MBTI.*

Кіріспе. Психологиялық типтерді зерттеу арқылы біз адамның таңдауы немесе ақпаратты қабылдау тәсілі бойынша әлеммен қалай байланыста болатындығын түсінуге тырысамыз. Адамның мінез-құлқын немесе Тұлғаны жеке сипаттамалармен өлшеуге болады. Тұлға психологиялық құрылым және оны жалпы мінез-құлық үлгісімен анықтауға болады[1]. Дәстүрлі әдіс бойынша тұлғаны талдау жазбаша тест, сұхбаттасу және сауалнама жүргізу сияқты көп уақытты қажет ететіні, қымбатқа түсетіні және пайдасыз нәтижеге әкелетіні байқалады. Адамдар әлеуметтік желілерді жиі қолданатындықтан, әлеуметтік желілерде олар жеке және мінез-құлық сипаттамаларымен бөліседі. Мұндай жеке ақпараттың үлкен көлеміне ойлар, сезімдер, ұмтылыстар, мақсаттар, сәтсіздіктер, қорқыныштар мен армандар, сондай-ақ өздеріне ұнайтын, ұнатпайтын дүниелері кіреді. Демек, тұлғаның типін анықтау үшін көп уақыт пен қаражатты қажет ететін әдістерді жеке адамның әлеуметтік желідегі жазбаларына сүйене отырып классификациялау тапсырмасына ауыстыруға болады[2,14]. Психологиялық типтерді біз әлеммен қарым-қатынасымызға байланысты туын-

*Адрес для переписки. E-mail: myngzhassar_akniyet@mail.ru

дайтын психологиялық сипаттамалар ретінде анықтауға болады және бұл көбінесе біздің оқуымызға және мамандық таңдауымызға әсер етеді. Сонымен қатар, осы ерекшеліктерді түсіну бізге конфликтілерді басқаруға, көшбасшылықты дамытуға, оқытуды жетілдіруге және басқа да көптеген дағдыларға көмектеседі.

Машиналық оқыту алгоритмдерін пайдаланушының әлеуметтік желі парақшасындағы мәліметтері арқылы оның мінез-құлқының типін шығаруға пайдалануға болады[15,16]. Бұл жұмыста MBTI моделіне негізделген психологиялық типтер ұсынылады.

MBTI адамдардың әлемді қалай қабылдайтынын және шешім қабылдайтындығын анықтау үшін қолданылады. Ол Юнг типологиясының негізінде пайда болған және АҚШ пен Еуропада кең таралған тұлға типологиясы[19]. Осы типологияның негізінде Myers-Briggs Type Index (MBTI) психологиялық тестілеу жүйесі құрылған болатын. Америка Құрама Штаттарында орта мектепті бітірушілердің 70% -ына дейін болашақ мамандығын таңдау мақсатында MBTI тұлғалық типін анықтау тестін тапсырады. Жыл сайын 2 миллионнан астам адам MBTI сауалнамасын толтырады. MBTI сауалнамасы 30 тілге аударылған және бүкіл әлемде қолданылады. Майерс-Бриггс типологиясын қолданудың негізгі қолданбалы бағыттарына өзін-өзі тану және жеке өсу, мансаптық өсу және кәсіптік бағдар беру, бизнес және консалтингте топ құру, клиенттермен өзара әрекеттесу және сату, адамның жеке тұлғасының жақсы жақтарына сәйкес келетін мамандық таңдау, жеке өмірде серіктес таңдау және отбасылық қатынастарды құруды атап айтып кетуге болады.

Әдеби талдау. Есептеу әдістері арқылы темпераменттің автоматты түрде танылуы көптеген бизнес секторлары мен әлеуметтік зерттеушілерге әлеуметтік медиа қолданушыларын түсінуге көмектеседі. Бүгінгі күнге дейін әдебиетте психологиялық типтерді жіктеуге байланысты бірнеше жұмыстар бар. Бұл бағыттағы жұмыстардың тапшылығының басты себебі – жіктеуіштерді оқытуға арналған деректерді іздеудің қиындығы.

Люккс пен Делеманс[3] голландтық «Жасанды өмір» деректі фильмі туралы студенттердің 145 эссесінен тұратын 200 000 сөздік Personae корпусын құрды. Сонымен қатар, студенттер MBTI профилін ұсынды. Бұл жұмыста авторлар авторлықтың атрибуциясын және тұлғаны болжауды орындады. Мәтіндік функцияларды шығару үшін жадқа негізделген беттік талдау (MBSP), n-грамм функциялары қолданылды. Тұлғаны болжау үшін, K-NN алгоритміне негізделген TiMBL (Memory Based Learning) әдісін қолдана отырып, 10 рет кросс-валидация жаттығуы өткізілді. Тәжірибелерде MBTI дихотомиясы үшін әрқайсысы 84 екілік классификациялық есептер болды. Авторлар интроверт-экстраверт және интуитивті қабылдаудың болжамдары жеткілікті дәл болды, орташа F-балл сәйкесінше 65,38% және 61,81% болды.

Комисин және Гуинн [4] Майерс-Бриггс моделі бойынша психологиялық типті анықтау үшін құжаттарды жіктеуге негізделген жүйе жасады. Өз тәжірибелерінде олар Naïve Bayes классификаторын және векторлық машиналарды қолданды. Деректер студенттердің MBTI және Best Best Future Self (BPFs) тесттерін орындаған студенттерге ұсынылатын жанжалдарды басқару бойынша аспирантураның бір бөлігі

ретінде жиналды. BPFS қазіргі және болашақтағы әр түрлі жағдайда өзін-өзі сипаттайтын элементтерден тұрады (мысалы, жұмыс, мектеп, отбасы, қаржы). Мәліметтер 2010-2011 жылдар аралығында үш семестрде жиналды. N-gram және Linguistic Inquiry and Word Count (LIWC) мәтіндерді бейнелеу үшін пайдаланылды.

Верховен және басқалар [5] алты мың тілде (голланд, неміс, француз, итальян, португал және испан) MBTI мәліметтер жиынтығын құрды, 18168 қолданушы және тілдерге таралған 34 миллионға жуық твиттер бар. Деректер базасын жасағаннан кейін авторлар гендерлік және Майерс-Бриггстің жеке параметрлерін (I / E, N / S, T / F және J / P) болжау үшін жіктеу тесттерін өткізді. Авторлар n-грамнан стандартты параметрлері бар LinearSVC қолданды.

MYERS - BRIGGS TYPE INDICATOR (MBTI). Майерс-Бриггс типологиясы - Юнг идеялары негізінде пайда болған тұлғаның жеке ерекшеліктерін анықтауға мүмкіндік беретін жүйе. Атақты психиатр және аналитикалық психологияның негізін қалаушы төрт негізгі психологиялық функция адамның қоршаған ортаны қабылдауына көмектеседі деп болжады [13,20]. Бұл функцияларға мыналар жатады: ойлау, сезім, интуиция және түйсік. MBTI жеке тұлғаны зерттеуге арналған 4 бинарлы шкаладан тұрады:

1. E - I шкаласы – энергетикалық бағдар:
 - E (Extraversion, экстраверсия) - сыртқы әлемге бағдар,
 - I (Introversion, интроверсия) - ішкі әлемге бағдар.
2. S - N шкаласы - ақпаратты қабылдау:
 - S (Sensing(ощущение), түйсіну) - фактілерге, егжей-тегжейлерге, ерекшеліктерге, статистикаға сүйену,
 - N (Intuition, интуиция) - мүмкіндіктерге, идеяларға, абстракцияға назар аудару.
3. T-F шкаласы - шешім қабылдау:
 - T (Thinking, ойлау) - объективті талдау басым,
 - F (Feeling, сезіну) – адами құндылықтар негізінде шешім қабылдау.
4. J - P шкаласы - әрекет:
 - J (Judging, төрелік) - ақпаратты жоспарлауға және алдын-ала тапсырыс беруге,
 - P (Perception, қабылдау) - жағдайларға сәйкес жоспарсыз әрекет етуді қалау.

Әр шкаладан тұлғаның жеке типтері таңдалыну арқылы 4 әріптен тұратын психотип анықталады. Мысалы, ISFJ типі:

- I - ішкі әлемге назар аударады;
- S - фактілер мен нақтылықты бағалайды;
- F - шешім қабылдауда өз сезімдеріне сенеді;
- J - құрылымдалған және жоспарланған өмірді дұрыс көреді.

Деректер сипаттамасы. Біз деректерімізді Kaggle-ден Myers-Briggs Personal Type (MBTI) деректер жиынтығынан алдық [6]. Деректер базасында екі бағаннан тұратын 8675 қатар бар. Әр қатарда екі түрлі деректер бар:

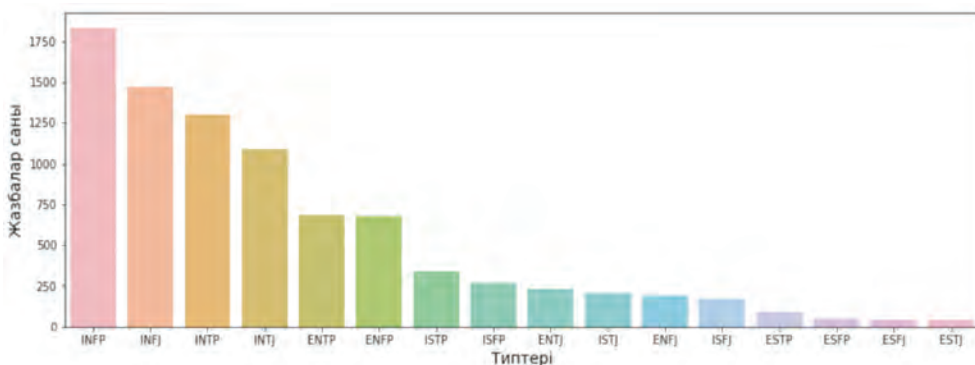
- Майерс-Бриггстің жеке типі (төрт әріптен тұратын код түрінде).
- Пайдаланушының твиттердегі соңғы 50 жазбасы (әр жазба «|||» арқылы бөлінген)

Төменде мәліметтер базасының алғашқы 5 жазбасының қысқартылған мысалы келтірілген:

	type	posts
0	ISFP	'Олар сандарсыз сурет салады Мен де istp д...
1	ISFP	'Рас, өкінішке орай, көпшілік оған қатысқысы к...
2	ISFP	1: Соғыс пен капитализм сияқты қажетсіз боктар...
3	ISFP	'рахмет, мен мұны қалай айтарымды білмеймін, б...
4	ISFP	'Жарайды, сондықтан мен көшіп келемін, ал мені...

1-сурет – Мәліметтер базасының алғашқы 5 жазбасы

Деректер жиынында нөлдік мәндер жоқ. Деректер жиынтығы біркелкі емес, яғни 16 тип әр түрлі үлестірілген (2-сурет). Мысалы, ең көп таралған INFP типі 1832 рет кездеседі, ал ең аз таралған ESTJ типі тек 39 рет кездеседі.



2-сурет – Деректердің таралуы

Мұндай біркелкі болмаудың себебі әлемде осы типтегі адамдардың біркелкі емес таралуы.

Зерттеу барысы. Деректер жиындығындағы деректерді 16 класқа бөліп, оларға машиналық оқыту алгоритмдерін қолданудың келесідей кемшіліктері бар:

Кластар арасында ұқсастықтар көп болуы мүмкін, мысалы INTJ және INTP типтері екі бөлек деп қарастырылса да, олардың айырмашылықтары айтарлықтай үлкен емес. Бұндай кластар арасында бірегейлікті анықтау қиынға түседі.

Әлеуметтік желілердегі тексттер бірыңғай емес және әр түрлі болғандықтан, біздің классификаторға өте ұқсас және шулы мәліметтер арасында айырмашылықтарды табуға тура келеді. Бұл нақты нәтижеге әкелмейді.

Сондықтан деректер жиынтығымен біркелкі жұмыс жасау үшін олар әр шкаладағы көрсеткіштері бойынша бір-біріне қарама-қарсы, өзара тәуелсіз екі бөлікке бөліп

алынды. Яғни адамның интроверсиясы / экстраверсиясы оның түйсінуімен / интуициясымен байланысты емес деп есептелді. Нәтижесінде 1-кестегідей деректерді аламыз.

1-кесте – Деректер жиынын әрбір жеке индикатор бойынша бөлу нәтижесі

Шкала №	МВТИ типі	Қатарлар саны	МВТИ типі	Қатарлар саны
1	Интроверсия, Introversion (I)	1999	Экстраверсия, Extroversion (E)	6676
2	Түйсіну, Sensing (S)	7478	Интуиция, Intuition (N)	1197
3	Ойлау, Thinking (T)	4694	Сезіну, Feeling (F)	3981
4	Төрелік, Judging (J)	5241	Қабылдау, Perceiving (P)	3434

Әр индикатор бойынша деректер теңдей бөлінбесе де, деректердің көлемі оқыту мен тестілеуге жетерлік. Егер де жоғарыдағы кестені диаграмма түсінде көрсетер болсақ:



1-диаграмма – Жеке индикаторлар бойынша деректердің таралуы

Деректерді оқытуға дайындау. Бұл деректер жиынтығындағы деректер твиттер қолданушыларының посттарынан алынғандықтан, онда сілтемелер, шылаулар, сандар, стоп-сөздер және т.б. қажет емес сөздер мен символдар бар болғандықтан, алдын ала деректерді тазалау жүргізілді. Сондай-ақ, сөздер векторларға айналдырылып, TF-IDF анықталды. Мәтінде кездесу жиілігі 10% -дан 70% -ға дейінгі сөздерді тану үшін sklearn кітапханасы пайдаланылды. Бірінші қадамда посттар токендерді санау матрицасына орналастырылды. Келесі қадамда модель лексика сөздігін үйреніп, термин-құжат матрицасын қайтарады. Содан кейін санау матрицасы нормализацияланған TF-IDF көрінісіне айналады, 797 сөз пайда болды.

Жіктеу тапсырмасы төрт бинарлы жіктеу тапсырмасына бөлінді, өйткені әрбір МВТИ типі төрт екілік класстан тұрады. Бұл қадамда әр типтің индикаторы жеке-жеке құрастырылды. Нәтижеде біз әр шкала бойынша бір индикатор таңдап алы-

нуы арқылы 4 әріптен тұратын тұлға типін аламыз. X айнымалысы мәтіндер үшін анықталған TF – ID-ке, ал Y айнымалысы MBTI типті екілік индикаторға пайдаланылды. MBTI типті индикаторлар жұбы жеке оқытылды, содан кейін мәліметтер sklearn кітапханасынан `train_test_split()` функциясы арқылы 80% оқыту және 20% тестілеу мәліметтер жиынтығына бөлінді.

Машиналық оқыту әдістері. Naive Bayes классификаторы – әр X және Y сипаттама жұптарын тәуелсіз деп қарастыратын Байес теоремасына негізделген оқытушымен оқыту алгоритмдеріне жататын әдіс. Naive Bayes классификаторлары өздерінің жеңіл болғандығына қарамастан, текстті классификациялау мен фильтрацияда жақсы жұмыс істейді. Бұл әдіс көп оқыту мен үлкен есептеу қуаттылығын қажет етпейді.

Boost – бұл нақты емес ережелерді біріктіру арқылы өте дәл болжам жасауға негізделген әдіс[10]. Extreme Gradient Boosting – бұл Gradient Boosting алгоритмін оңтайландырылған түрі[11].

Екі әдісті жүзеге асыру үшін Numpy, NaiveBayesClassifier, XGBoost және sklearn қолданылды. Деректер жиынтығының 80%-ы оқыту деректері, ал қалған 20%-ы тесттік деректерге бөлініп оқыту жүргізілді. Тесттік деректер жиынтығы үшін екі әдістің дәлдігі анықталды.

Зерттеу нәтижелері. 2-кестеде әр жеке индикаторлар жұбы үшін тесттік жиынтықта екі классификатордың көрсеткен дәлдіктері:

2-кесте – Тесттік жиынтықтағы нәтижелер

MBTI индикаторлары	Naive Bayes	XGBoost
Introversion (I) / Extroversion (E)	76.14%	76.32%
Intuition (N) – Sensing (S)	86.69%	86.10%
Feeling (F) - Thinking (T)	68.13%	69.54%
Judging (J) – Perceiving (P)	59.60%	59.10%

Introversion (I) / Extroversion (E), Intuition (N) – Sensing (S), Judging (J) – Perceiving (P) жұптары үшін Naive Bayes классификаторы жақсы нәтиже көрсетті, тек Feeling (F) - Thinking (T) үшін ғана XGBoost аз ғана жоғары дәлдікті көрсетті. Екі әдістің де дәлдіктерінде аса қатты айырмашылық байқалмады.

Қорытынды. Бұл зерттеуде MBTI индивидуалды типі негізінде тұлғаның типін болжау процесін автоматтандыруға арналған машиналық оқытудың Naive Bayes және XGBoost әдістері зерттелді. Табиғи тілді өңдеу құралы (NLTK) және python-дағы pandas, numpy, sklearn, re, matplotlib, NaiveBayesClassifier, XGBoost қолданылды. Naive Bayes және XGBoost моделдерінің дәлдігі бағаланды және екі классификатор да бірдей дерлік нәтиже көрсетті. Қорыта келе, қазақ тілі үшін қолдануда қарапайым, әрі көп есептеу қуаттылығын қажет етпейтін, туралайтын параметрлері аз Naive Bayes классификаторын пайдаланған тиімдірек. Осы зерттеуде келтірілген әдістеде қазақ тілі үшін дәлдік пен сенімділік жақсы деңгейде болды. Бұл жұмыс NLP практиктері

мен психологтарына тұлғаның типтерін және онымен байланысты когнитивті процестерді анықтауда тиімді көмек көрсете алады.

ӘДЕБИЕТ

1 М. Стивенсон, Введение в нейролингвистическое программирование. [M. Stivenson, Vvedenie v nejrolingvisticheskoe programmirovanie]

2 Tieger P.D.; Barron-Tieger, B. Do What You Are: Discover the Perfect Career for You through the Secrets of Personality Type, 4th ed.; Sphere: London, UK, 2007.

3 Luycx K. Daelemans W. Personae: корпус для предсказания автора и личности на основе текста. Конференция по языковым ресурсам и оценке (LREC). 2008 г. [Luycx K, Daelemans W. Personae: korpus dlya predskazaniya avtora i lichnosti na osnove teksta. Konferenciya po yazyk-ovym resursam i ocenke (LREC). 2008 g.]

4 Барбуто Дж. Э. Младший. Критика индикатора типа Майерс-Бриггс и его операционализация психологических типов Карла Юнга. Психологические отчеты. С. 611–625, 1997. [Barbutto Dzh. E. Mladshij. Kritika indikatora tipa Majers-Briggs i ego operacionalizaciya psiholog-icheskikh tipov Karla Yunga. Psihologicheskie otchety. S. 611–25, 1997.]

5 Верховен Б., Далеманс В., Планк Б. TwiSty: Многоязычный корпус по стилометрии Twitter для определения пола и личности. LREC. 2016 г. [Verhoven B., Daelemans V., Plank B. TwiSty: Mnogoyazychnyj korpus po stilometrii Twitter dlya opredeleniya pola i lichnosti. LREC. 2016 g.]

6 MBTI kaggle деректер жиынтығы: <https://www.kaggle.com/datasnaek/mbti-type>. [MBTI kaggle derekter zhiyntygy: <https://www.kaggle.com/datasnaek/mbti-type>.]

7 Lu J., Xiong C., Parikh D., Socher R. Knowing When to Look: Adaptive Attention via a Visual Sentinel for Image Captioning. 2017. IEEE Computer Vision and Pattern Recognition.

8 М. Комисин, Curry Guinn. Identifying personality types using document classification methods. Proceedings of the 25th International Florida Artificial Intelligence Research Society Conference, FLAIRS-25, pages 232–237, 01 2012.

9 Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. Пособие/Большакова Е.И., Клышинский Э.С., Ландэ Д.В., Носков А.А., Пескова О.В., Ягунова Е.В. – М.: МИЭМ, 2011. – 272 с. [Avtomaticeskaya obrabotka tekstov na estestvennom yazyke i komp'yuternaya lingvistika: ucheb. Posobie / Bol'shakova E.I., Klyshinskij E.S., Lande D.V., Noskov A.A., Peskova O.V., YAgunova E.V. – М.: МИЭМ, 2011. – 272 s.]

10 Freund Y.; Schapire R.E. A decision-theoretic generalization of on-line learning and an application to boosting. J. Comput. Syst. Sci. 1997, 55, 119–139

11 Friedman J.H. Greedy function approximation: A gradient boosting machine. Ann. Stat. 2001, 1189–1232.]

12 Грант С. Ингерсолл, Томас С. Мортон, Эндрю Л. Фэррис. Обработка неструктурированных тестов. Поиск, организация и манипулирование. / Пер. с англ. Слинкин А.А. – М.: ДМК Пресс, 2015. – 414 с. [Grant S. Ingersoll, Tomas S. Morton, Endryu L. Ferris Obrabotka nestrukturirovannyh testov. Poisk, organizaciya i manipulirovanie. / Per. s angl. Slinkin A.A. – М.: DMK Press, 2015. – 414 s..]

13 Rick Harrington, Donald A. Loffredo, MBTI personality type and other factors that relate to preference for online versus face-to-face instruction 2010 y.

14 Navonil Majumder, Soujanya Poria, Alexander Gelbukh, Erik Cambria, “Deep learning-based document modeling for personality detection from text,” published by the IEEE Computer Society, IEEE Intelligent Systems 2017.

15 Sang-Woon Kim and Joon Min Gil, “Classification systems based on TFIDF and LDA schemes,” in Human Centric Computing and Information Sciences, 30, 2019.

16 P.S. Dandannavar, S.R. Mangalwede, P.M. Kulkarni, “Social Media Text – A Source for Personality Prediction”, International Conference on Computational Techniques, Electronics and Mechanical Systems, IEEE, 2018.

17 A.Helen Victoria, M . Vijayalakshmi. An outcome based Comparative study of different text classification Algorithms, International Journal of Pure and Applied Sciences, Mathematics, ijpm.eu, 2018.

18 Dhore M L, Varpe K M, “Emotion Detection based on Images and Captions on Social Media” in International Conference on Analytical Innovation (ICAI) 2020, In press.

19 J.W. Pennebaker and L.A. King, “Linguistic Styles: Language Use as an Individual Difference,” J. Personality and Social Psychology, vol. 77, no. 6, 1999, pp. 1296 –1312.

20 Rayne Hernandez, Ian Scott Knight, “Predicting Myers-Briggs Type Indicator with Text Classification”, 31st Conference on Neural Information Processing Systems, NIPS, 2017.

**А. А. МЫНЖАСАР, А. Б. КУЛЬДЖАБЕКОВ,
Б. С. ДАРИБАЕВ, А. Н. ТЕМИРБЕКОВ**

Казахский национальный университет имени аль-Фараби, г. Алматы, Казахстан

ИСПОЛЬЗОВАНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОПРЕДЕЛЕНИЯ MYERS-BRIGGS TYPE INDEX (MBTI) ТИПОВ ЛЮДЕЙ

Статья посвящена проблеме машинного обучения в области компьютерной лингвистики, в частности идентификации психологических типов людей на основе текстовых сообщений в социальных сетях. Целью данной статьи является изучение методов машинного обучения Naive bayes и Extreme Gradient Boosting (XGBoost) для создания классификатора для казахского языка, определяющего тип индекса типа Майерса-Бриггса (MBTI) на основе текстовых образцов публикаций людей в социальных сетях. Представлен ход исследовательских экспериментов с использованием методов машинного обучения и результаты исследования, а также сопоставлены полученные результаты.

Ключевые слова: Naive Bayes, XGBoost, TF-IDF, MBTI.

**A. A. MYNGZHASSAR, A. B. KULDZHABEKOV,
B. S. DARIBAYEV, A. N. TEMIRBEKOV**

Al-Farabi Kazakh National University, Almaty, Kazakhstan

USING MACHINE LEARNING METHODS TO DETERMINE MYERS-BRIGGS TYPE INDEX (MBTI) TYPES OF PEOPLE

The article is based on the problems of machine learning in the field of computer linguistics, in particular, the identification of psychological types of people on the basis of text messages on social networks. The purpose of this article is to study the methods of machine learning Naive bayes and Extreme Gradient Boosting (XGBoost) to create a classifier for the Kazakh language, which determines the type of Myers-Briggs Type Index (MBTI) based on text samples of people's posts on social networks. The course of research experiments in the use of machine learning methods and the results of the study are presented and the results obtained are compared.

Keywords: Naive Bayes, XGBoost, TF-IDF, MBTI.