

Однако эти конструкции не смогли обеспечить два ценнейших свойства балластного слоя: его упругость и способность распределять нагрузки от подвижного состава, передаваемые на балласт через рельсы и шпалы. Поэтому жесткие основания, уложенные непосредственно под рельсы, могут иметь лишь ограниченные сферы применения (мосты, эстакады, тоннели). Опыт эксплуатации железнодорожных путей с асфальтовым основанием показал, что применение слоя асфальта под балластным слоем уменьшает расходы на текущее содержание пути.

ЛИТЕРАТУРА

- [1] Железнодорожный путь /Т.Г. Яковлева, Н.И. Карпушенко, С.И. Клинов, Н.Н. Путря, М.П. Смирнов: Под ред. Т.Г. Яковлевой 2-е изд. М.: «Транспорт», 2004. -407 с.
- [2] СТП 091.56.010-2005. Текущие содержание железнодорожного пути. Технические требования и организация работ. Белорусская железная работа. Минск. -2006. - 284 с.
- [3] Расчеты и проектирование железнодорожного пути: Учебное пособие для студентов вузов ж. -д. трансп./ В.В. Виноградова, А.М. Никонов, Т.Г. Яковлева и др.; Под ред. В.В. Виноградова и А.М. Никонова. - М.: Маршрут, 2003. - 486 с.
- [4] Н.В. Бубликов, П.В. Ковтун, А.М. Патласов, А.Г. Жуковец. Проектирование рельсовой колеи. - Гомель. БелГУТ. 1996. - 53 с.
- [5] Елсаков Н.Н. Практическое руководство по текущему содержанию стрелочных переводов. - М.: Транспорт 2000.- 107 с.
- [6] Устройство и эксплуатация железнодорожного пути: Пособие/ В.И. Матвеев, П.В. Ковтун, А.Г. Жуковец, Т.И. Есева - Гомель: БелГУТ, 2004. -114 с.
- [7] Матвеев В.И., Илютин В.И., Есева Т.И. Требования к оформлению дипломных и курсовых проектов. - Гомель. БелГУТ. 2003. - 69 с.
- [8] Расчет и проектирование основных параметров рельсовой колеи - Методические указания к курсовому проекту для студентов ??? - ?V курсов специальности 270204 - «Строительство железных дорог, путь и путевое хозяйство» часть I, Екатеринбург, УрГУПС, 44 стр.
- [9] Расчеты и проектирование железнодорожного пути: Учебное пособие для студентов вузов ж. -д. трансп./ Под ред. В.В. Виноградова и А.М. Никонова. - М.: Маршрут, 2003.--486 с.
- [10] Яковлева Т.Г., Шульга В.Я., Амелин С.В. и др. Основы устройства и расчет ж.д. пути / Под ред. С.В.Амелина и Т.Г. Яковлевой. - М.: Транспорт, 1990.

Байкенжева А.С.

Қазақстан Республикасының теміржол көлігінің жол төселімін тиімділігі мен қауіпсіздігін қамтамасыз етудің онтайлы жағдайы мен тәсілдері.

Түйіндеме. Мақалалар жалпы темір жол балластының негізгі құрылымдары туралы әдеби талдаулар жүргізіліп, ғылыми түсініктерге, нақты тәжірибелерге сілтеме жасалған. Темір жол саласы бойынша жол төселімінің тұрақтылығын көтеру мақсатында бірнеше материалдармен жасалған жұмыстар ұсынылып, жұмыс барысында қарастырылып отырған Запшта – Зырянск жол бөлігіндегі сын көтермейтін жол бөліктері туралы зерттеулер жүргізіліп, бірнеше статистикалық мәліметтер айнаған.

Түйін сөздер: темір жол, жол төселімі, демормация, балласт төңірегі

ӨЖ 81.93.29

M.A. Bolatbek, Sh.Zh. Mussiraliyeva

DETECTION OF EXTREMIST TEXTS USING MACHINE LEARNING METHODS

Abstract. A sharp increase in the number of materials with extremist content on the Internet poses a threat to national security of the country. Unfortunately, terrorist and extremist organizations began to spread information in the Kazakh language too. In this regard, in order to prevent citizens to join to extremist organizations it becomes urgent to monitor the users' activity on the Internet. In this paper, authors analyzed the Ansar1 forum to determine the content of extremist texts in English and used machine learning methods to determine the accuracy of the relation of input text to an extremist class. To perform the classification, a corpora was constructed which consists of the texts of above mentioned forum and machine learning methods such as support vector machine, naive Bayes classifier, logistic regression, classification trees and random forest classifier were used.

Key words. Extremism, text classification, machine learning methods, naive Bayes classifier, logistic regression, support vector machine.

М.А. Болатбек, Ш.Ж. Мусиралиева

(ал-Фараби атындағы Қазақ Ұлттық Университеті, Алматы, Қазақстан Республикасы

E-mail: {bolatbek.milana, mussiraliyevash}@gmail.com)

## ЭКСТРЕМИСТИК МӘТІНДЕРДІ МАШИНАЛЫҚ ОҚЫТУ ӘДІСТЕРІ АРҚЫЛЫ АНЫҚТАУ

**Аннотация.** Интернет желісіндегі экстремистік материалдарды қамтитын сайттар санының күрт артуы ұлттық қауіпсіздік саласы үшін белгілі бір қауіп төндіреді. Өкінішке орай, террористік және экстремистік топтар қазақ тілінде де ақпарат тарата бастаған. Осы орайда азаматтардың лаңкестік топтарға кіруінің алдын алу үшін қолданушылардың Интернет желісіндегі әрекеттерін бақылау маңызды болып табылады. Бұл жұмыста авторлар ағылшын тіліндегі экстремистік мәтіндердің мазмұнын анықтау мақсатында Ansa1 форумына талдау жасап, машиналық оқыту әдістері көмегімен кіріс мәтіннің экстремистік топқа жату дәлдігін анықтайды. Жіктеу үшін аталған форум мәтіндерінен тұратын корпус құрылып, тірек векторлар машинасы, аңқау Байес жіктеуіші, логистикалық регрессия, жіктеу ағаштары және кездейсоқ орман жіктеуіші сияқты машиналық оқыту әдістері пайдаланылды.

**Кілтсөздер.** Экстремизм, мәтінді жіктеу, машиналық оқыту әдістері, аңқау Байес жіктеуіші, логистикалық регрессия, тірек векторлар машинасы.

### Кіріспе

Интернет желісі мен ақпараттық технологиялардың қарқынды дамуы ұлттық қауіпсіздік саласына айтарлықтай қауіп төндіреді. Себебі соңғы онжылдық ішінде белгілі экстремистік ұйымдар өз қызметтерін Интернет кеңістігінде белсенді жүргізетін болған. Интернетте 10000-ға жуық экстремистік ұйым сайттары тіркелген [1]. Экстремистік ұйымдар әлеуметтік желілерді, әр түрлі форумдарды өз топтарына жана мүшелер шақыру, оларға насихат жүргізу немесе лаңкестік әрекеттерді орындауға қажетті қаражат жинау және т.с.с. үшін пайдаланады.

Террористік ұйымдар қазақ тілінде де экстремизм мен терроризмді насихаттайтын материалдар тарата бастаған. Ал мұндай мазмұндағы материалдардың саны ғаламторда геометриялық прогрессиямен өсуде. 2016 жылы 12 экстремистік акцияның жолы кесілді. 182 азамат террористік бап бойынша заң алдында жауап берді. Терроризмді қаржыландырғаны үшін халықаралық террористік ұйымдардың 3 көмекшісі қылмыстық жауапкершілікке тартылды. 5 экстремистік топ мүшесі шетелдік әріптестерімізге жіберілді. ҰҚК ақпараттық қолдауының арқасында шетелдік арнайы органдар арқылы халықаралық террористік ұйымдардың 14 мүшесі ұсталған.

Қазақ тілінде экстремизм мен терроризм идеясын насихаттайтын жеке материалдардың саны күрт артуда. Бұл экстремистік топтардың арнайы аудиторияда жұмыс істей бастағандығын көрсетеді. Олар What's app, Viber, Twitter және т.б. әлеуметтік желілері арқылы жұмысын жандандыруда [2]. 2017 жылы Дін істері және азаматтық қоғам министрлігінің мамандары ғаламтордағы 5527 ақпарат ресурстарына зерттеу жүргізген. Зерттеу қорытындыларына сәйкес, олардың ішінде 2487 заңсыз материалдардың бар екендігі анықталды. 2017 жыл басынан бері теріс діни контенті бар 2,5 мың сайт анықталып, бұғатталған [3]. Осы себепті, экстремистік мазмұндағы сайттарды анықтау мақсатында Интернет ресурстарды автоматты түрде бақылау жүйелерін құру өте маңызды болып табылады.

### Әдебиеттерге шолу

Экстремистік мазмұндағы мәтіндерді анықтау үшін ғалымдар бірнеше әдістерді ұсынады. Әлеуметтік желілердегі экстремистік ұйымдардың жұмыстарын анықтау мен болжау әдістері [4] жұмысында ұсынылады. Зерттеушілер форумдағы мәтіндерді жіктеу үшін аңқау Байес үлгілері, логистикалық регрессия, жіктеу ағаштары, тірек векторлары машинасы әдістерін пайдаланады.

Ғалымдар [5] жұмыста Твиттер әлеуметтік желісіндегі радикалды контентті автоматты түрде анықтауға ұмтылады. Жіктеу үшін ағылшын тіліндегі сөздерге 1 (негативті) және 3 (позитивті) аралығындағы белгілі бір ұпайларды тағайындайтын Dictionary of Affect in Language (DAL) сияқты сөздіктер пайдаланылады. Авторлар тәжірибе жүргізу барысында Support Vector Machine (SVM), аңқау Байес және AdaBoost сияқты әр түрлі жіктеушітерді пайдаланған.

Қолданушылардың Твиттер платформасында жариялайтын твиттері арқылы олардың терроризмге итермелейтін әрекеттерін анықтау әдісі [6] жұмысында көрсетілген. Мәліметтер Twitter API көмегімен белгілі бір кілтсөздерді беру (мысалы, "террорист") арқылы жинақталады. Келесі кезеңде жинақталған мәліметтер барлық тэгтерді, қажетсіз ақпаратты өшіру арқылы тазартылады. Содан кейін мәліметтер аңқау Байес үлгісі көмегімен оң, теріс және бейтарап санаттарға жіктеледі.

Ғалымдар экстремистік топтар жні пайдаланатын сөздерді анықтау үшін [7] жұмыста Part of Speech Tagging әдісін пайдаланады. Аталған әдіс жні пайдаланатын сөздерді анықтап, оларды сөз

таптары, мысалы зат есім, етістік, сын есім және т.с.с. бойынша топтастырады. Бұл әдіс кілтсөздер ретінде белгілі бір сөз таптарын ғана қарастыру кезінде ыңғайлы болып келеді.

Авторлар [8] жұмыста веб-ресурстардағы қазақ тіліндегі экстремистік мәтіндерді анықтауға қатысты зерттеу жүргізген. Аталған есепті шешу үшін бес кезеңнен тұратын үлгі құрылған. Жұмыстың мақсаты экстремистік топтар жиі қолданатын кілтсөздерді анықтау болып табылады, анықталған кілтсөздер әрі қарай мәтіндерді “экстремистік” және “бейтарап” санаттарға бөлу үшін пайдаланылады. Зерттеу жұмысын жүргізу үшін қазақ тіліндегі экстремистік мәтіндер корпусы құрылған. Корпуса барлығы 150 мәтін бар, олардың 80-і экстремистік сипатта болса, қалғаны жаңалық сайттарынан және т.б. жиналған бейтарап мәтіндер болып табылады. Кілтсөздерді анықтау үшін TF-IDF әдісі пайдаланылады. Анықталған кілтсөздер мәліметтер қорына түрлі морфологиялық нұсқаларымен бірге енгізілген. Нәтижесінде берілген кіріс мәтін ішінен мәліметтер қорындағы экстремистік кілтсөздерді іздейтін бағдарлама құрылған.

#### Материалдар мен әдістер

Бұл жұмыста авторлар Dark Web Portal Project [9] жобасы барысында жинақталған форумдардың біріне талдау жасайды. Dark Web Portal Project жобасы жеті тілдегі әр түрлі 28 форум мәліметтерін қамтиды. Аталған форумдарда экстремистік және жалпы діни мәселелер талқыланады, олардың бірқатары радикалдық ислами сипаттағы мәтіндер болып келеді. Авторлар аталған форумдардың ішіндегі Ansa1 форумына тоқталды. Ол жерде барлығы 29942 хабарлама бар. Хабарламалар 12/2008-1/2010 мерзімі аралығында жинақталған. Қатысушылардың жалпы саны - 382. Негізгі тілі ағылшын тілі болса да, арасында араб, орыс, түрік және кавказ ұлттарының тіліндегі хабарламалар кездеседі. Жіктеу дәлдігін арттыру мақсатында авторлар келесі тапсырмаларды орындады:

1. Ansa1 форумындағы барлық 29492 хабарламаны оқылды.
2. Хабарламаларды оқи отырып, экстремистік мазмұндағы хабарламалардың екі жеке корпусы құрылды, олардың бірі машиналық оқыту әдістерін оқыту үшін, ал екіншісі аталған әдістер арқылы жіктеу жүргізу үшін пайдаланылады.
3. Корпусағы мәтіндердің орфографиялық қателері түзетілді, барлық тыныс белгілері мен сілтемелер өшірілді.
4. Корпусағы мәтіндерге Портер стемминг алгоритмі қолданылды, яғни мәтіндегі әр сөздің жалғаулары мен жұрнақтары алынып тастап, тек негізі ғана қалдырылады. Бұл сөздің әр түрлі қосымша жалғанған барлық нұсқаларын бір сөз ретінде қарастыру үшін жасалды.
5. Нәтижесінде корпусағы мәтіндердің жалпы саны 417, олардың 246-сы оқыту корпусында, ал қалған 171-і тәжірибе жүргізу корпусына орналастырылды.
6. Келесі кезеңде жинақталған мәліметтерге Python Scikit-Learn оқыту кітапханасында жіктеу жүргізілді. Жіктеу үшін тірек векторлар машинасы (SVM), аңқау Байес жіктеуші (naive Bayes classifier), логистикалық регрессия (logistic regression), жіктеу ағаштары (classification trees) және кездейсоқ орман жіктеуші (random forest classifier) машиналық оқыту әдістері пайдаланылды.

#### Тірек векторлар машинасы

Тірек векторлар машинасы (ағылшынша *SVM, support vector machine*) әдісінің негізгі идеясы оқытушы жиындағы оң және теріс мысалдарды максималды түрде ажырататын шешімдер беті болатын гиперкөлемді құру болып табылады. Тірек векторларды оқыту алгоритмін құру тірек вектор мен кіріс көлемінен алынған вектордың скаляр көбейтіндісінің ядросы түсінігіне негізделеді. Тірек векторлар оқытушы таңдаманың ішкі жиыны болып табылады. Әдістің орындалуы барысында екі сынып шекарасындағы бейнелерге (мысалдарға) іздеу жүргізіледі. Мұндай бейнелер тірек векторлар болып табылады. Олар мәліметтердің шешімдер көлеміне (бөлінетін екі сынып шекаралары арасындағы аймақ) ең жақын орналасқан нүктелері болып табылады. Жалпы алғанда тірек векторлар машинасы әдісі көмегімен жіктеу дегеніміз бастапқы бейнелер жиынын екі сыныпқа бөлетін және бір сынып бейнелері үшін нөлден кіші, ал басқа сынып векторлары үшін нөлден үлкен мәнге ие болатын қандай да бір сызықты функцияны іздеу болып табылады. Жіктеудің ең жақсы функциясы күтілетін тәуекелі ең кіші функция болып табылады. Күтілетін тәуекел — бұл жіктеу қателігінің күтілетін деңгейі [10].

#### Аңқау Байес жіктеуші

Аңқау Байес сипаттамалар арасындағы тәуелсіздікті болжайтын Байес теоремасын қолдануға негізделетін ықтималдық жіктеуші. Аңқау Байес  $d$  құжатының  $c$  классына тиісті болуының  $p$  ықтималдығын:  $p(c|d)$  есептейді. Жіктеуге арналған  $d$  құжаты  $d = (d_1, \dots, d_n)$  векторы арқылы бейнеленеді, шартты ықтималдықты Байес теоремасын пайдалану арқылы келесі түрде жазуға болады [5]:

$$p(c|d) = \frac{p(c)p(d|c)}{p(d)}$$

*Логистикалық регрессия*

Логистикалық регрессия — бұл тәуелді айнымалы екі санаты бар санаттық айнымалы, ал тәуелсіз айнымалылар сандық немесе санаттық айнымалылар болғанда қолданылатын болжау үлгісінің бір түрі. Сонымен қатар, ол таңдалған кездейсоқ оқиғаның орындалу немесе орындалмау ықтималдығын да бағалайды. Бұл үлгі оқиғаның белгілі бір санатқа жататынын немесе жатпайтынын бағалай отырып, оны жіктей алады.

logit — айнымалының логарифм коэффициенттеріне тең функция. Егер  $p = Y=1$  (оқиғаның орындалу) ықтималдығы болса, онда  $\frac{p}{1-p}$  коэффициенттерге сәйкес келеді.  $p$  ықтималдығының logit функциясы келесі түрде жазылады:

$$\text{logit}(p) = \log \frac{p}{1-p}$$

Логистикалық регрессия үлгісі:

$$\text{logit}(p) = b_0 + b_1 * X_1 + b_2 * X_2 + \dots + b_k * X_k$$

$X_1$  — тәуелсіз айнымалылар мәні,  $b_1$  — есептелуі логистикалық регрессия тапсырмасы болып табылатын коэффициенттер,  $b_0$  — белгілі бір тұрақты шама.

Логистикалық регрессияда тәуелді айнымалы шын мәнінде logit болады, ол коэффициенттер логарифмі болып табылады:

$$\text{logit}(p) = \ln \frac{p}{1-p}$$

Сонымен ізделіп отырған ықтималдық келесідей болады [11]:

$$p = \frac{e^{\text{logit}(p)}}{1 + e^{\text{logit}(p)}}$$

*Жіктеу ағаштары*

Жіктеу ағашы (шешім қабылдау ағаштары немесе регрессиялық ағаш деп те аталуы мүмкін) — жіктеуіштерді, сонымен қатар регрессия үлгілерін бейнелеу үшін пайдаланылатын болжау үлгісі. Ағаш “жапырақтар” мен “бұталардан” тұрады. Шешім ағашының “бұталарына” мақсат функция тәуелді болатын атрибуттар, “жапырақтарына” мақсат функцияның мәндері, ал қалған түйіндерге оқиғалар ажыратылатын атрибуттар жазылады. Жаңа оқиғаны жіктеу үшін ағаш бойымен жапыраққа дейін түсіп, сәйкес мәнді шығару керек. Операцияларды зерттеу кезінде шешімдер ағашында шешімдердің сатылы үлгісі мен олардың салдарлары қарастырылады. Шешім қабылдаушы көшбасшы шешімдер ағашын өз мақсатына жетуге мүмкін болатын стратегияны таңдау үшін қолданады. Жіктеу ағаштары нысанды немесе үлгіні олардың атрибуттарының мәндері негізінде алдын ала анықталған сыныптар жиынына жіктеу үшін пайдаланылады [12].

*Кездейсоқ орман жіктеуіші*

Кездейсоқ орман жіктеуіші — жіктеу және регрессия тапсырмалары үшін жиі пайдаланылатын дәлдігі жоғары бейсызық алгоритм. Ол шешімдер ағашына негізделеді. Қандай да бір қабылдануы тиіс шешім бар деп есептейік. Аталған шешімді қабылдауға көмектесетін айнымалылар бар, олардың кейбірінің салмақтары басқаларына қарағанда жоғары болып келеді. Шешімдер ағашының алгоритмі компьютерге оның оқытушы жиынындағы қай айнымалылардың ең маңызды екендігін анықтауға мүмкіндік береді. Ол аталған айнымалыларды ағаштың ең жоғарғы тұсына орналастырады да, маңыздылығы аз айнымалыларға біртіндеп ауысады. Мұндай әдіс айнымалыларды топтастыруға мүмкіндік береді. Өкінішке орай, шешімдер ағашының жоғарғы деңгейлері жауапқа қатты әсер етеді және егер жаңа мәліметтер оқытушы таңдаманың үлестіріміне сай келмесе, үлгінің жалпылау қабілеті төмендейді. Дәл осы жерде “кездейсоқ орман” алгоритмі негізіндегі үлгі көмекке келеді. Ағаштар жиынын құру аталған тәуекелді азайтады. Жіктеу барысында жауап іздеуде жауаптардың көбі, ал регрессия жағдайында орташасы таңдалады. Дауыстар мен орташаларды пайдалану арқасында басқа алгоритмдер бере бермейтін толық ықтималдықтарды алу мүмкіндігі туындайды [13].

*Нәтижелер және талдау*

Құрылған корпустың 417 мәтіннің 171-іне Python Scikit-Learn оқыту кітапханасында жіктеу жүргізілді. Жіктеу үшін тірек векторлар машинасы (SVM), аңқау Байес жіктеуіші (naive Bayes classifier), логистикалық регрессия (logistic regression), жіктеу ағаштары (classification trees) және кездейсоқ орман жіктеуіші (random forest classifier) машиналық оқыту әдістері пайдаланылды. Жіктеу нәтижесінде берілген

кіріс корпустағы мәтіндердің экстремистік топқа жату дәлдігі анықталды. Жіктеу нәтижелері 1-кестеде келтірілген. Ең жоғары дәлдік аңқау Байес жіктеуіші пайдаланған кезде алынды (0.90).

1-кесте. Кіріс мәтіндерді жіктеу нәтижелері

Машиналық оқыту әдісі	Дәлдік
Тірек векторлар машинасы	0.69
Аңқау Байес жіктеуіші	0.90
Логистикалық регрессия	0.73
Жіктеу ағаштары	0.73
Кездейсоқ орман жіктеуіші	0.88

Кіріс мәтіндердің экстремистік топқа жату дәлдігімен қатар Ansar1 форумындағы ең жиі кездесетін терминдер де анықталды. Экстремистік мазмұндағы мәтіндерде жиі кездесетін терминдер тізімі 2-кестеде келтірілген.

2-кесте. Ағылшын тіліндегі экстремистік мәтіндерде жиі кездесетін терминдер

Термин	Жиілігі	Термин	Жиілігі
allah	152	destroy	24
mujahideen	60	enemi	22
inshaallah	50	war	10
kill	36	punish	10
attack	34	mujahedin	7
fight	31	allaah	7
kuffar	30	money	5
inshallah	25	inschallah	5

Зерттеу жұмысы барысында анықталғандай, берілген форумдағы ағылшын тіліндегі жиі кездесетін бірқатар діни терминдер [8] жұмыста анықталған қазақ тіліндегі кілтсөздермен сәйкес келеді екен. Қазақ және ағылшын тілдеріндегі экстремистік мәтіндерде жиі кездесетін діни терминдер тізімі 3-кестеде келтірілген.

3-кесте. Қазақ және ағылшын тілдеріндегі хабарламалардағы діни терминдер сәйкестігі

Қазақ тіліндегі кілтсөздер	Ағылшын тіліндегі кілтсөздер
аллах/алла/аллах	allah/allaah
соғыс/согыс	war
кәпір/кафир	kuffar
иншааллах/иншаллах/иншаллах	inshallah/inschallah
мужахид/можахид/муджахед/муджахид	mujahideen/mujahedin

#### Қорытынды

Берілген жұмыста ағылшын тіліндегі Ansar1 форумындағы мәтіндерге талдау жасалып, оларға машиналық оқыту әдістері көмегімен жіктеу жүргізілді. Жіктеу үшін Python Scikit-Learn оқыту кітапханасында тірек векторлар машинасы (SVM), аңқау Байес жіктеуіші (naive Bayes classifier), логистикалық регрессия (logistic regression), жіктеу ағаштары (classification trees) және кездейсоқ орман жіктеуіші (random forest classifier) машиналық оқыту әдістері пайдаланылды. Жіктеу нәтижесі бойынша ең жоғары дәлдік аңқау Байес жіктеуіші пайдаланған кезде алынды. Келешекте машиналық оқыту әдістерінің дәлдігін арттыру мақсатында жіктеуді бірнеше ережелерді қатар пайдалану арқылы жүргізу жоспарлануда.

#### ӘДЕБИЕТТЕР

[1] Жаворонкова Т.В. Террористік және экстремистік ұйымдардың Интернет желісін пайдалануы // Оренбург мемлекеттік Университеті хабаршысы. — 2015. — №3(178). — 30 б.

- [2] Қазақ тілінде «сойқан» сайттар көбейіп кетті. <http://jasqazaq.kz/2017/01/31/aza-tilinde-soj-an-sajttar-k-bejir-ketti/>. Қаралған күні: 10.03.2018.
- [3] Қазақстанда жыл басынан бері экстремизмге үгіттейтін 2,5 мың сайт анықталған. <https://sn.kz/kk/sn-zap-zaman/26169-aza-standa-zhyl-basynan-beri-ekstremizmge-gittejtin-2-5-my-sajt-any-tal-an>. Қаралған күні: 15.03.2018.
- [4] Дж.Р.Сканлон. Экстремистік кибер-шақыртуды автоматты түрде анықтау және болжау // Магистрлік диссертация. — Виргиния Университеті, 2014. — 5-7 б.
- [5] Е.Омер. Твиттердегі жиһаттық хабарламаларды анықтау үшін машиналық оқытуды пайдалану // Магистрлік диссертация. — Уппсала Университеті, 2015. — 5-14 б.
- [6] С.А. Азизақ, И.А.Азиз. Терроризмді машиналық оқытуды пайдалану арқылы сентимент талдау негізінде анықтау // Инженерия және қолданбалы ғылымдар журналы. — 2017. — №12(3). — 691-698 б.
- [7] Р.Скривенс, Р.Франк. Интернеттегі радикалдық мәтіндерді сентимент негізінде жіктеу // Еуропалық зерде және ақпараттық қауіпсіздік конференциясы. — Уппсала, 2016. — 104-107 б.
- [8] М.А.Болатбек, Ш.Ж.Мусиралиева, У.А.Тукаев. Қазақ тілді веб-контенттегі экстремистік бағытты анықтауға арналған түйінсөздер мәліметтер қорын құру // Қазақ Университеті. Математика, механика, информатика. — 2018. — №1(97). — 134-142 б.
- [9] Көлеңкелі Веб және геосаяси веб-зерттеулер. <https://ai.arizona.edu/research/dark-web-geo-web>. Қаралған күні: 05.11.2017.
- [10] О.В.Китова, И.Б.Колмаков, И.А.Пеньков. Инвестиция көрсеткіштерін болжауға арналған тірек векторлар машинасы әдісі // Статистика және экономика. — 2016. — №4. — 27-29 б.
- [11] В.С.Рават. Логистикалық регрессия. — Лакшмибай физикалық ғылымдар ұлттық институты, 2018. — 2-5 б.
- [12] О.З.Маймон, Л.Рокач. Шешімдер ағашымен білім жинау: Теория және қосымшалар. — Екінші басылым. — АКШ: World Scientific Publishing, 2015. — 304 б.
- [13] Х.Бринк, Д.Ричардс, М.Феверолф. Машиналық оқыту. . — М: Питер, 2017. — 336 б.

Болатбек М.А., Мусиралиева Ш.Ж.

Определение экстремистских текстов с помощью методов машинного обучения

Резюме. В данной работе авторы для определения содержания экстремистских текстов на английском языке анализировали форум Ansa1 и с помощью методов машинного обучения определили точность отнесения входного текста в экстремистский класс. Для выполнения классификации был построен корпус, который состоит из текстов вышеуказанного форума и были использованы методы машинного обучения такие как, машина опорных векторов, классификатор наивного Байеса, логистическая регрессия, деревья классификации и классификатор случайного леса.

Ключевые слова. Экстремизм, классификация текста, методы машинного обучения, классификатор наивного Байеса, логистическая регрессия, машина опорных векторов.

Bolatbek M.A., Mussiraliyeva Sh.Zh.

Detection of Extremist Texts Using Machine Learning Methods

Summary. In this paper, authors analyzed the Ansa1 forum to determine the content of extremist texts in English and used machine learning methods to determine the accuracy of the relation of input text to an extremist class. To perform the classification, a corpora was constructed which consists of the texts of above mentioned forum and machine learning methods such as support vector machine, naive Bayes classifier, logistic regression, classification trees and random forest classifier were used.

Key words. Extremism, text classification, machine learning methods, naive Bayes classifier, logistic regression, support vector machine.

УДК 004.8

Н.П. Азанов, Ж.Б. Манас

## THE SECURITY PROBLEMS AND CRYPTOGRAPHIC STRENGTH OF ELLIPTIC CURVES

Abstract: Every day we see an increase in computing power and at the same time an increase in the number of attempts to hack cryptosystems in order to obtain any data. Today, information technology is developing very rapidly at the same time as the value of all kinds of digital information is increasing. With the growth of value at the same time increases the cost of information security [2].

All information processed on computer networks and the Internet must be well protected. This fact requires continuous improvement of information security mechanisms. It is for this reason that the creation and use of secure data encryption systems is currently important.

Key words: cryptography, elliptic curve, singularity, supersingularity.

