ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

ACADEMIC SCIENTIFIC JOURNAL OF COMPUTER SCIENCE

№2
2026

ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)

CENTRAL ASIAN ACADEMIC
RESEARCH CENTER

ACADEMIC SCIENTIFIC JOURNAL OF COMPUTER SCIENCE

2 (358)

APRIL – JUNE 2026

PUBLISHED SINCE JANUARY 1963
PUBLISHED 4 TIMES A YEAR

ALMATY, NAS RK

Academic Scientific Journal of Computer Science
ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)
Owner: «Central Asian Academic Research Center» LLP (Almaty).
Certificate **№ KZ77VPY00121154** on the re-registration of the periodical printed and online publication of the information agency, issued on **05.06.2025** by the Republican State Institution «Information Committee» of the Ministry of Culture and Information of the Republic of Kazakhstan
Subject area*: information and communication technologies.*
Currently: *included in the list of journals recommended by the CCSES MSHE RK in the direction of «Information and communication technologies».*
Periodicity: *4 times a year.*
http://www.physico-mathematical.kz/index.php/en/

Academic Scientific Journal of Computer Science
ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)
Меншіктеуші: «Орталық Азия академиялық ғылыми орталығы» ЖШС (Алматы).
Ақпарат агенттігінің мерзімді баспасөз басылымын, ақпарат агенттігін және желілік басылымды қайта есепке қою туралы ҚР Мәдениет және Ақпарат министрлігі «Ақпарат комитеті» Республикалық мемлекеттік мекемесі **05.06.2025** ж. берген **№ KZ77VPY00121154** Куәлік.
Тақырыптық бағыты: *ақпараттық-коммуникациялық технологиялар*
Қазіргі уақытта: *«ақпараттық-коммуникациялық технологиялар» бағыты бойынша ҚР БҒМ БҒСБК ұсынған журналдар тізіміне енді.*
Мерзімділігі: *жылына 4 рет.*
http://www.physico-mathematical.kz/index.php/en/

Academic Scientific Journal of Computer Science
ISSN 2518-1726 (Online),
ISSN 1991-346X (Print)
Собственник: *ТОО «Центрально-азиатский академический научный центр» (г. Алматы).*
Свидетельство о постановке на переучет периодического печатного издания, информационного агентства и сетевого издания **№ KZ77VPY00121154.** Дата выдачи **05.06.2025**
Тематическая направленность: *информационно-коммуникационные технологии.*
В настоящее время: *вошел в список журналов, рекомендованных КОКСНВО МНВО РК по направлению «информационно-коммуникационные технологии».*
Периодичность: *4 раза в год.*
http://www.physico-mathematical.kz/index.php/en/

CONTENTS

COMPUTER SCIENCE

МАЗМҰНЫ

КОМПЬЮТЕРЛІК ҒЫЛЫМДАР

СОДЕРЖАНИЕ

КОМПЬЮТЕРНЫЕ НАУКИ

ACADEMIC SCIENTIFIC JOURNAL OF COMPUTER SCIENCE
ISSN 1991-346X
Volume 2.
Namber 358 (2026). 286–301

https://doi.org/10.32014/2026.2518-1726.439

IRSTI: 28.23.37; 20.23.17
UDC 004.912; 004.89

© Marassulov U.A.*, Kazbekova G., 2026.
Khoja Akhmet Yassawi International Kazakh-Turkish University,
Turkistan, Kazakhstan.
E-mail: marasulov.usen2024@ayu.edu.kz

TF-IDF-BASED FAKE NEWS DETECTION IN KAZAKH AND RUSSIAN

Marassulov Ussen — PhD student, Khoja Akhmet Yassawi International Kazakh-Turkish University, Turkistan, Kazakhstan,
E-mail: marasulov.usen2024@ayu.edu.kz, ORCID ID: https://orcid.org/0009-0008-0801-1229;
Kazbekova Gulnur — Candidate of Technical Sciences, Associate Professor, Khoja Akhmet Yassawi International Kazakh-Turkish University, Turkistan, Kazakhstan,
E-mail: gulnur.kazbekova@ayu.edu.kz, ORCID ID: https://orcid.org/0000-0002-2756-7926.

Abstract. Relevance: The rapid spread of unverified information in digital media and social networks makes automatic fake news detection an important task for natural language processing and machine learning. For Kazakh, this task is particularly relevant because open labeled corpora are limited and ready-made models are not sufficiently adapted to local media discourse. Purpose: The study evaluates TF-IDF-based machine learning models for fake/real text classification in Kazakh and Russian and forms an interpretable baseline for later comparison with transformer-based approaches. Methods: The experiments used a balanced Kazakh-Russian corpus of 1,808 texts, with 452 fake and 452 real texts in each language. Logistic Regression, Linear SVM and Complement Naive Bayes were tested with word-level and character-level TF-IDF features in bilingual, monolingual and cross-lingual scenarios. Model quality was assessed using accuracy, precision, recall, F1-score and confusion matrices. Results: In bilingual, Kazakh-only and Russian-only settings, the best models achieved Macro-F1 of approximately 0.985. Cross-lingual evaluation showed clear directional asymmetry: Kazakh-to-Russian transfer reached Macro-F1 = 0.654, whereas Russian-to-Kazakh transfer reached Macro-F1 = 0.926. Practical significance: The results provide a reproducible and explainable benchmark for Kazakh-Russian fake news detection. They also define a basis for future comparison with multilingual BERT and XLM-RoBERTa under stricter source-based and temporal evaluation settings.

Keywords: fake news, disinformation, Kazakh language, Russian language, TF-IDF, machine learning, cross-lingual classification

For citations: Marassulov U.A., Kazbekova G. TF-IDF-based fake news detection in Kazakh and Russian. Academic Scientific Journal of Computer Science, 2026. — No.2. – P. 286–301. DOI https://doi.org/10.32014/2026.2518-1726.439

© Марасулов У.А.*, Казбекова Г., 2026.
Қожа Ахмет Ясауи атындағы Халықаралық қазақ-түрік университеті,
Түркістан, Қазақстан.
E-mail: marasulov.usen2024@ayu.edu.kz

ҚАЗАҚ ЖӘНЕ ОРЫС ТІЛДЕРІНДЕГІ ЖАЛҒАН ЖАҢАЛЫҚТАРДЫ TF-IDF АРҚЫЛЫ АНЫҚТАУ

Марасулов Усен — PhD докторант, Қожа Ахмет Ясауи атындағы Халықаралық қазақ-түрік университеті, Түркістан, Қазақстан,
E-mail: marasulov.usen2024@ayu.edu.kz, ORCID ID: https://orcid.org/0009-0008-0801-1229;
Казбекова Гулнур — техника ғылымдарының кандидаты, қауымдастырылған профессор, Қожа Ахмет Ясауи атындағы Халықаралық қазақ-түрік университеті, Түркістан, Қазақстан,
E-mail: gulnur.kazbekova@ayu.edu.kz, ORCID ID: https://orcid.org/0000-0002-2756-7926.

Аннотация. Өзектілігі: Цифрлық медиа мен әлеуметтік желілерде тексерілмеген ақпараттың жылдам таралуы жалған жаңалықтарды автоматты түрде анықтауды табиғи тілді өңдеу мен машиналық оқытудың маңызды қолданбалы міндетіне айналдырды. Қазақ тілі үшін бұл мәселе ерекше өзекті, себебі ашық белгіленген корпустар аз, ал дайын модельдер жергілікті медиамәтіндердің тілдік және тақырыптық ерекшеліктеріне толық бейімделмеген. Мақсаты: Зерттеудің мақсаты - қазақ және орыс тілдеріндегі fake/real мәтіндерді жіктеу үшін TF-IDF белгілеріне негізделген машиналық оқыту модельдерін бағалау және кейінгі трансформерлік модельдермен салыстыруға арналған түсіндірілетін baseline қалыптастыру. Әдістері: Экспериментте 1808 мәтіннен тұратын теңгерілген қазақ-орыс корпусы қолданылды: әр тілде 452 fake және 452 real мәтін қамтылды. Logistic Regression, Linear SVM және Complement Naive Bayes модельдері word-level және character-level TF-IDF белгілерімен екітілді, біртілді және кросс-тілдік сценарийлерде салыстырылды. Бағалау accuracy, precision, recall, F1-score және confusion matrix көрсеткіштері арқылы жүргізілді. Негізгі нәтижелер: Екітілді, қазақша және орысша бөлек бағалау режимдерінде ең жақсы модельдер шамамен Macro-F1 = 0.985 нәтижесін көрсетті. Кросс-тілдік бағалау бағытқа тәуелді айырмашылықты анықтады: қазақ тілінен орыс тіліне тасымалдауда Macro-F1 = 0.654, ал орыс тілінен қазақ тіліне тасымалдауда Macro-F1 = 0.926 болды. Практикалық маңызы: Нәтижелер қазақ-орыс медиакеңістігінде жалған жаңалықтарды анықтауға арналған қайталанатын

әрі түсіндірілетін бастапқы өлшем береді және multilingual BERT пен XLM-RoBERTa модельдерімен кейінгі салыстыруларға негіз болады. Бұл өлшем болашақ эксперименттердің сапасын салыстырмалы түрде бағалауға мүмкіндік береді.

Түйін сөздер: жалған жаңалықтар, дезинформация, қазақ тілі, орыс тілі, TF-IDF, машиналық оқыту, кросс-тілдік классификация

© Марасулов У.А.*, Казбекова Г., 2026.
Международный казахско-турецкий университет имени Ходжи Ахмеда Ясави, Туркестан, Казахстан.
E-mail: marasulov.usen2024@ayu.edu.kz

ВЫЯВЛЕНИЕ ЛОЖНЫХ НОВОСТЕЙ НА КАЗАХСКОМ И РУССКОМ ЯЗЫКАХ TF-IDF-МОДЕЛЯМИ

Марасулов Усен — PhD-докторант, Международный казахско-турецкий университет имени Ходжи Ахмеда Ясави, Туркестан, Казахстан,
E-mail: marasulov.usen2024@ayu.edu.kz, https://orcid.org/0009-0008-0801-1229;
Казбекова Гулнур — кандидат технических наук, ассоциированный профессор, Международный казахско-турецкий университет имени Ходжи Ахмеда Ясави, Туркестан, Казахстан,
E-mail: gulnur.kazbekova@ayu.edu.kz, https://orcid.org/0000-0002-2756-7926.

Аннотация. *Актуальность.* Быстрое распространение непроверенной информации в цифровых медиа и социальных сетях делает автоматическое выявление ложных новостей значимой задачей обработки естественного языка и машинного обучения. Для казахского языка данная проблема особенно актуальна, поскольку открытых размеченных корпусов немного, а готовые модели недостаточно адаптированы к локальному медиадискурсу. *Цель.* Оценить модели машинного обучения на основе TF-IDF для классификации fake/real текстов на казахском и русском языках, а также сформировать объяснимый baseline для последующего сравнения с трансформерными моделями. *Методы.* В эксперименте использован сбалансированный казахско-русский корпус из 1808 текстов: по 452 fake и 452 real текста на каждом языке. Модели Logistic Regression, Linear SVM и Complement Naive Bayes сравнивались с word-level и character-level TF-IDF-признаками в билингвальном, одноязычных и кросс-языковых сценариях. Качество классификации оценивалось по показателям accuracy, precision, recall, F1-score и confusion matrix, что позволило учесть не только общий результат, но и направление ошибок. *Результаты и выводы.* В билингвальном, казахском и русском режимах лучшие модели достигли значения Macro-F1 около 0,985. Кросс-языковая оценка выявила асимметрию переноса: обучение на казахском языке и тестирование на русском дало Macro-F1 = 0,654, тогда как обучение на русском языке и тестирование на казахском достигло Macro-F1 = 0,926. Полученные результаты формируют воспроизводимый и

интерпретируемый ориентир для выявления ложных новостей в казахско-русском медиапространстве и создают основу для дальнейшего сравнения с multilingual BERT и XLM-RoBERTa. Практическая значимость исследования заключается в возможности использования предложенного подхода как базовой модели для мониторинга дезинформации, анализа медиаконтента и разработки интеллектуальных инструментов проверки достоверности текстов на казахском и русском языках.

Ключевые слова: ложные новости, дезинформация, казахский язык, русский язык, TF-IDF, машинное обучение, кросс-языковая классификация, fake news, Macro-F1

Introduction. Online media and social platforms have made the movement of news almost immediate. This speed improves access to information, yet it also helps unverified, distorted or intentionally misleading texts circulate widely. Such content may influence public trust, political and economic choices, and health-related decisions. For that reason, automatic fake news detection is now an applied research task with direct relevance for natural language processing, machine learning and information security. Existing studies mainly address high-resource languages, especially English, where labeled corpora, fact-checking platforms, pretrained models and benchmark datasets are more accessible. Kazakh differs from this setting. Open labeled data are limited, the language has rich morphology, and local misinformation practices are not fully reflected in English-language resources. This gap is consistent with reviews that discuss dataset quality, model limitations and the lack of multilingual and cross-lingual datasets (Harris et al., 2024: 222; Alghamdi et al., 2026: 353), with work on multi-level Kazakh-Russian annotation (Sambetbayeva et al., 2025: 215), and with an early benchmark for Kazakh fake news detection (Telman et al., 2026: 708).

Kazakhstan's information environment is bilingual: Kazakh and Russian are used side by side in news production and consumption. Therefore, fake news detection should not be treated only as a single-language task. It is important to test whether one model can handle both languages in a shared setting and whether signals learned from one language remain useful for fake/real classification in the other. For low-resource languages, such testing is especially valuable because it shows how far a better represented language can support classification when labeled data are scarce.

The purpose of this paper is to estimate the initial capacity of classical machine learning models for Kazakh-Russian fake/real classification and to report the result as an explainable baseline. Two research questions guide the study: how effective are TF-IDF features with linear classifiers on a bilingual Kazakh-Russian corpus, and can a model trained in one language distinguish fake and real texts in another language? The aim is not to present a completed fact-checking product, but to define a transparent reference point for later comparison with transformer-based approaches.

The contribution of the study is fourfold. It uses a balanced experimental dataset containing fake and real texts in Kazakh and Russian. It compares several TF-IDF baselines under the same protocol in monolingual, bilingual and cross-lingual regimes. It also shows that transfer between the two languages is directional: Russian-to-Kazakh transfer performs better than Kazakh-to-Russian transfer. Finally, the interpretation of results explicitly takes into account source bias, topic bias, temporal bias and near-duplicate risks.

Literature review. Automatic fake news detection is not limited to a standard text classification problem. News content, user reactions, propagation networks and source credibility may all affect the decision boundary. Shu et al. (2017: 22) draw attention to news content, user engagement and social context, while Zhou and Zafarani (2020: 109) describe fake news signals through writing style, propagation patterns, false knowledge and source credibility. Benchmark research also shows that model performance depends on the dataset, label quality and evaluation scenario, not only on the selected algorithm (Galli et al., 2022: 237).

Work on the diffusion of misinformation confirms the practical importance of this field. Vosoughi, Roy and Aral (2018: 1146) found on Twitter data that false news may spread faster and more widely than true news. From this perspective, detection models are connected not only with technical classification, but also with information security, media trust and public resilience.

The quality of the dataset is one of the main conditions for reliable evaluation. LIAR supported the analysis of short political statements through multi-level truthfulness labels (Wang, 2017: 422). FEVER connected claim verification with evidence extraction (Thorne et al., 2018: 809), and FakeNewsNet supplemented news texts with social context and spatiotemporal information (Shu et al., 2020: 171). X-Fact extended this direction to multilingual fact-checking (Gupta and Srikumar, 2021: 675). These resources demonstrate that annotation design and corpus structure determine how model scores should be understood.

A separate issue is that high scores may reflect stylistic or source-related differences instead of factual reasoning. Hamed et al. (2023: e20382) connect fake news detection quality with dataset size, label quality, feature representation and data fusion. Thibault et al. (2025: 5801) show that spurious correlations and label noise can reduce generalization. For this reason, the TF-IDF results in the present study are interpreted as evidence of distinguishable signals inside the corpus, not as evidence that factual verification has been solved.

In multilingual settings, evidence-based verification becomes even more important. Dementieva and Panchenko (2021: 310) showed that evidence in another language may improve monolingual fake news detection. The Multiverse study continues this idea by showing that cross-language comparison can serve as an additional explainable signal for classification (Dementieva et al., 2023: 77). This direction fits the Kazakhstani media context, where the same event may be reported in Kazakh and Russian with different wording, source choices and emphases.

The interpretation of evaluation results also depends on data completeness and

cleanliness. Galli et al. compared traditional machine learning and deep learning approaches in a fake news benchmark (Galli et al., 2022: 237). FakeNewsNet combines several information dimensions (Shu et al., 2020: 171), whereas Thibault et al. emphasize the control of label quality, leakage and spurious correlations (Thibault et al., 2025: 5801). Accordingly, the metrics reported in this paper should be treated as baseline measurements that require further checking with source-based, topic-based and temporal splits.

Low-resource and multilingual conditions make these problems harder. Recent reviews name dataset bias, model constraints, limited multilingual data and weak cross-lingual generalization among the key obstacles for fake news detection systems (Harris et al., 2024: 222; Alghamdi et al., 2026: 353). Cross-lingual and cross-domain transfer learning points to the possibility of transferring knowledge from high-resource to lower-resource settings (Providel et al., 2025: 287). De et al. also report the usefulness of multilingual BERT for low-resource fake news classification (De et al., 2021: 9). These findings support the need for carefully designed corpora and evaluation scenarios for Kazakh.

Multilingual transformer models provide an important comparison line. BERT introduced contextual text encoding that improved many NLP tasks (Devlin et al., 2019: 4171), while XLM-RoBERTa achieved strong cross-lingual transfer through large-scale multilingual pretraining (Conneau et al., 2020: 8440). For Kazakh and Russian, Sambetbayeva et al. propose moving beyond a binary fake/real label toward multi-level annotation, including CLAIM, SOURCE, EVIDENCE, DISINFORMATION_TECHNIQUE, AUTHOR_INTENT and TARGET_AUDIENCE (Sambetbayeva et al., 2025: 215). Telman et al. compared TF-IDF, a translation-based cross-lingual approach and XLM-RoBERTa for Kazakh fake news detection (Telman et al., 2026: 708). The present article therefore defines the TF-IDF baseline rather than replacing transformer models.

Overall, the literature shows that a reproducible and transparent baseline for Kazakh-Russian fake/real classification is still necessary. More complex deep learning models can be judged fairly only after the initial level of data quality, source bias and cross-lingual transfer has been measured. This study addresses that preliminary step.

Materials and main methods. Dataset. The experiments were conducted on a bilingual corpus of fake and real texts in Kazakh and Russian. The texts were collected from open sources, cleaned and deduplicated. After deduplication, the complete dataset contained 2,740 records: 452 fake and 518 real Kazakh texts, and 895 fake and 875 real Russian texts. The sources included fact-checking resources, materials documenting misinformation and mainstream news portals. The source_file field was kept for internal control but was not used as an input feature.

To limit the influence of language and class imbalance, the experimental subset was balanced by selecting 452 texts for each language-class pair. As a result, the final corpus used in the experiments contained 1,808 texts. This design makes model comparison clearer, but it does not reproduce the real distribution of fake

and real content in the media environment. Therefore, the reported values describe performance under controlled laboratory conditions.

Table 1 – Balanced dataset composition by language and class

Language	Fake texts	Real texts	Total
Kazakh	452	452	904
Russian	452	452	904
Total	904	904	1808
Note: Compiled by the authors based on the balanced experimental dataset.			

The Fake class was formed from fact-checking and misinformation-related sources, including factcheck_kz_zhalgan, gov_factcheck_fake_claims, nofake_new_wp_fake_kk_ru and provereno_media_fake_ru. The Real class was compiled from mainstream news sources such as Egemen, Kazinform, Informburo, Tengrinews, Zakon and user-provided Kazakh news. In this study, the real label indicates the origin of the text from a trusted news source; it does not mean that each item was independently fact-checked. Thus, the experiment measures textual fake/real discrimination rather than absolute truth verification. Model input combined the title, claim and main text fields. Metadata such as ID, source_file, URL and date were not provided to the classifier. Empty texts, invalid labels and records with incorrect language values were removed before training. This choice reduces direct source memorization, although indirect source cues may still appear through style, topic, text structure or portal-specific wording.

Deduplication was based on the title, claim and main text fields. Exact duplicate entries were removed, and records with conflicting labels were excluded from the experiment. This step lowers the risk of train/test leakage. Still, lightly edited texts or paraphrases describing the same event may remain in the data; therefore, external testing is needed before extending the conclusions to the wider information space. The two classes were also considered from the viewpoint of label origin. Fake texts often come from refutation or fact-checking-style materials, where denial, explanation and rejection of a source can be frequent. Real texts more often follow editorial news style, official information patterns and portal-specific formatting. These differences provide useful classification signals, but they may also lead the model to learn genre and source features rather than factual truth itself.

Experimental scenarios

Five experimental scenarios were used to evaluate the models.

Table 2 – Description of experimental scenarios

Scenario	Description
bilingual	Joint training and testing on Kazakh and Russian texts
kk_only	Training and testing only on Kazakh texts
ru_only	Training and testing only on Russian texts
cross_kk_to_ru	Training on the Kazakh train split and testing on all Russian texts
cross_ru_to_kk	Training on the Russian train split and testing on all Kazakh texts
Note: Compiled by the authors based on the experimental protocol.	

In the bilingual scenario, the train, validation and test split was stratified by both language and label. In kk_only and ru_only, stratification was based on the label. The approximate proportions were 70% for training, 15% for validation and 15% for testing. In cross-lingual scenarios, the model was trained on the training part of one language and evaluated on all 904 texts of the other language. Random seed = 42 was fixed throughout. Cleaning and balancing were completed before splitting; the validation part was retained for protocol consistency rather than broad hyperparameter search.

Features and models

TF-IDF was used to represent each text as a vector. Two feature configurations were tested. The first used word-level unigrams and bigrams; the second used character-level 3-5 n-grams. Character n-grams are relevant for Kazakh and Russian because they can capture inflectional endings, suffixes, recurring short patterns and orthographic similarity.

The comparison included five baseline configurations.

Table 3 – Compared TF-IDF baseline models

Model name	Feature type	Classifier
word_tfidf_logreg	Word TF-IDF, 1-2 n-grams	Logistic Regression
char_tfidf_logreg	Character TF-IDF, 3-5 n-grams	Logistic Regression
word_tfidf_svm	Word TF-IDF, 1-2 n-grams	Linear SVM
char_tfidf_svm	Character TF-IDF, 3-5 n-grams	Linear SVM
word_tfidf_cnb	Word TF-IDF, 1-2 n-grams	Complement Naive Bayes
Note: Compiled by the authors based on the experimental protocol.		

For Logistic Regression and Linear SVM, class_weight = balanced was applied. The word-level TF-IDF vocabulary was limited to 100,000 features, and the character-level vocabulary to 120,000 features. The minimum document frequency was min_df = 2; sublinear_tf was applied for Logistic Regression and Linear SVM. All experiments were implemented in scikit-learn. The vectorizer and classifier were fitted inside a single Pipeline, so the vocabulary and TF-IDF weights were learned only from the training split; test texts were not involved in feature construction.

TF-IDF was chosen as a baseline because it is simple, fast and relatively interpretable, which is useful in a low-resource setting. At the same time, it does not place Kazakh and Russian words into a shared semantic space. For that reason, cross-lingual performance may depend on shared topics, style, symbols or class-specific lexical patterns rather than on semantic understanding. This limitation is central to the interpretation of the experiment.

Evaluation metrics

Model quality was assessed with Accuracy, Precision, Recall and F1-score. Macro-F1 was used as the main comparison metric because it gives the fake and real classes equal weight. This is important in cross-lingual evaluation, where a model may overpredict one class while Accuracy hides the imbalance. Confusion matrices were added to show the direction and concentration of errors.

Reproducibility and bias control

Reproducibility was supported by fixing seed = 42 for data splitting and stochastic model components. All models were tested under the same scenarios and with the same textual fields. This design improves comparability, although it cannot fully remove topic and source differences that may exist within the corpus.

Three types of bias were considered during interpretation. Source bias may occur because fake and real texts come from different categories of websites. Topic bias appears when the two classes cover different subject areas. Temporal bias can inflate expected generalization when training and test texts are close in time. For these reasons, the high scores are reported as controlled baseline results, not as production-level accuracy.

Results

Table 4 summarizes the aggregated results. Figures 1-3 illustrate the scenario-level Macro-F1 and Accuracy distributions, and Figure 4 shows class-level F1 values. The best performance was observed in monolingual and bilingual regimes. In the bilingual scenario, word_tfidf_svm and char_tfidf_svm produced the same result: Accuracy = 0.9853 and Macro-F1 = 0.9853. In the Kazakh-only setting, word_tfidf_logreg and word_tfidf_svm reached Macro-F1 = 0.9853. In the Russian-only setting, word_tfidf_svm produced the highest score.

Table 4 – Best results by experimental scenario

Scenario	Best model	Accuracy	Macro-F1
bilingual	word_tfidf_svm	0.9853	0.9853
kk_only	word_tfidf_logreg	0.9853	0.9853
ru_only	word_tfidf_svm	0.9853	0.9853
cross_kk_to_ru	word_tfidf_cnb	0.6869	0.6540
cross_ru_to_kk	word_tfidf_cnb	0.9259	0.9257
Note: Compiled by the authors based on the experimental results.			

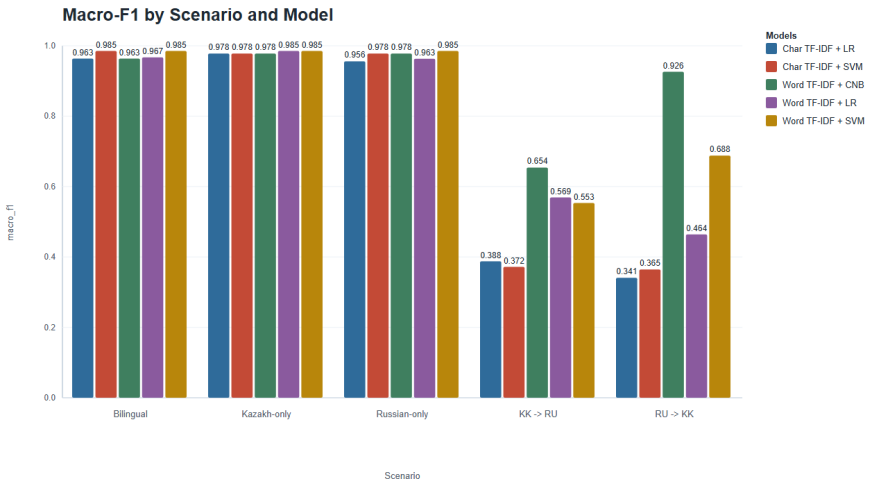

Figure 1 – Macro-F1 scores by experimental scenario

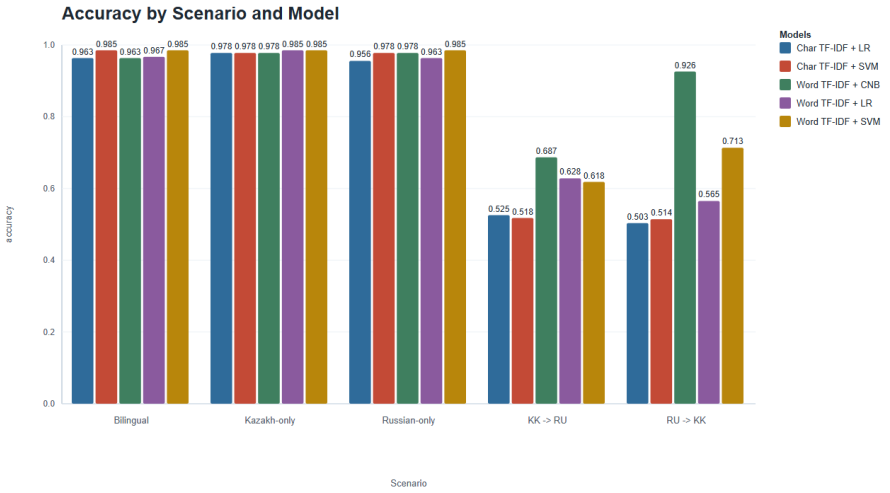

Figure 2 – Accuracy scores by experimental scenario

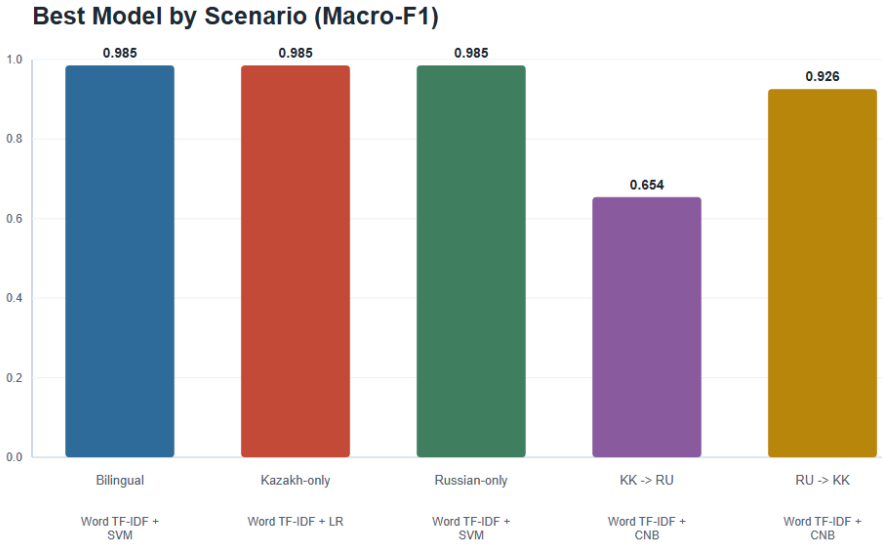

Figure 3 – Macro-F1 scores of the best models in each scenario

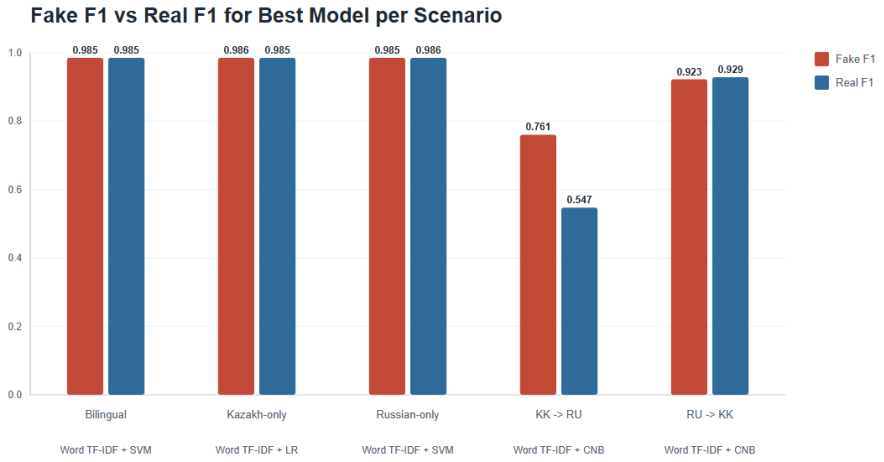

Figure 4 – F1 scores for fake and real classes in the best models

Figure 5 presents the confusion matrix for the bilingual scenario. The best model correctly classified 268 of 272 test texts. In the Fake class, 134 texts were detected correctly and 2 were assigned to the Real class. In the Real class, 134 texts were detected correctly and 2 were assigned to the Fake class. The errors are balanced across the two classes, so no clear class bias is observed in this scenario.

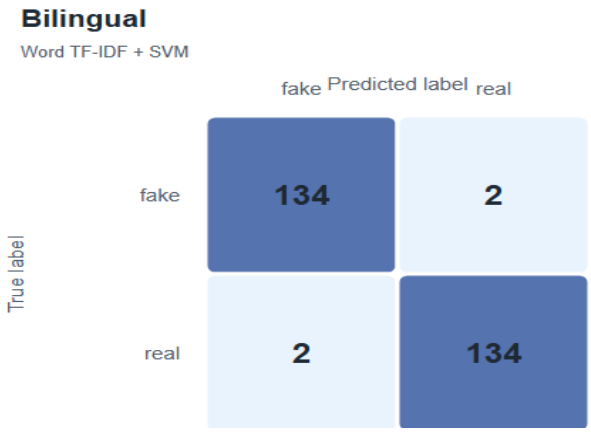

Figure 5 – Confusion matrix of the best model in the bilingual scenario

In the Kazakh-only scenario, the model classified 134 of 136 test texts correctly. All fake texts were detected, while 2 real texts were classified as fake. In the Russian-only scenario, the total number of correct predictions was also 134 of 136, but the error pattern was different: all real texts were retained, and 2 fake texts were classified as real.

Cross-lingual performance was strongly directional. In cross_kk_to_ru, the model was trained on the Kazakh training split and tested on Russian texts. The

best result was obtained by word_tfidf_cnb, with Accuracy = 0.6869 and Macro-F1 = 0.6540. The confusion matrix indicates that fake Russian texts were detected well, but real Russian texts were often moved to the fake class: 450 of 452 fake texts were correct, whereas 281 of 452 real texts were misclassified as fake. Figure 6 shows the corresponding matrix.

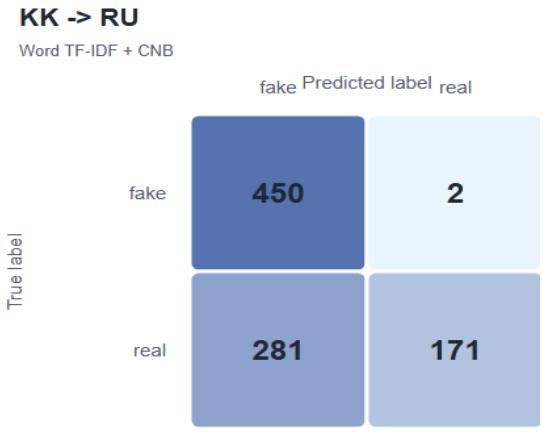

Figure 6 – Confusion matrix of the model trained on Kazakh and tested on Russian

The reverse direction produced a more stable result. In cross_ru_to_kk, the model was trained on the Russian training split and tested on Kazakh texts. The best model was again word_tfidf_cnb, reaching Accuracy = 0.9259 and Macro-F1 = 0.9257. In total, 837 of 904 test texts were classified correctly; this includes 399 of 452 fake Kazakh texts and 438 of 452 real Kazakh texts. Figure 7 presents the corresponding confusion matrix.

Figure 7 – Confusion matrix of the model trained on Russian and tested on Kazakh

Error analysis

The confusion matrices show where the main errors occur. In monolingual and bilingual regimes, the number of errors is small and the distribution between classes is almost even. In cross_kk_to_ru, the main weakness is the assignment of real Russian texts to the fake class: 281 of 452 real texts were classified as fake. This may mean that fake-related signals learned from Kazakh data overlap with stylistic or topical patterns in real Russian news.

In cross_ru_to_kk, the error distribution is less extreme: 399 of 452 fake texts and 438 of 452 real texts were recognized correctly. One possible explanation is that the Russian training set covers broader fake/real distinctions that are also present in the Kazakh test set. This explanation remains provisional and should be checked with source-based splitting, topic-controlled splitting and manual analysis of misclassified texts.

Discussion. The results indicate that classical TF-IDF models form a strong baseline for fake/real classification in this corpus. Macro-F1 values close to 0.985 in monolingual and bilingual scenarios show that the Kazakh and Russian texts contain clear class-related lexical signals. These signals may come from topic selection, writing style, source format, claim structure or wording typical of fact-checking materials. Thus, the high scores should be understood as evidence of separable textual patterns, not as proof of deep factual understanding.

The bilingual result suggests that Kazakh and Russian texts can be handled within one classification space. This is relevant for Kazakhstan's bilingual information environment. Nevertheless, TF-IDF does not encode contextual meaning; it relies on word and n-gram distributions. Therefore, the result shows that the dataset contains shared lexical and stylistic cues, but it does not prove that the model captured the full semantics of both languages.

The main analytical result is the asymmetry of cross-lingual transfer. Training on Kazakh led to weaker performance on Russian texts, while training on Russian transferred more successfully to Kazakh texts. Several explanations are possible: the Russian training set may be wider in topic and style, the source structure of the Kazakh data may differ from the Russian test set, or Complement Naive Bayes may preserve some lexical likelihood patterns better in cross-lingual settings. These are hypotheses for future ablation and source/topic analysis, not confirmed causal explanations.

This asymmetry requires cautious interpretation. TF-IDF does not build a shared semantic space for Kazakh and Russian, nor does it identify translation equivalents. It works with symbolic, stylistic and topical similarities present in the data. Therefore, the strong cross_ru_to_kk result is best understood as a combination of shared dataset signals and statistical adaptation rather than genuine semantic transfer.

The practical implication has two parts. First, TF-IDF should remain in future experiments as a simple, fast and interpretable comparison point for Kazakh-Russian fake/real classification. Second, multilingual transformer models,

including multilingual BERT and XLM-RoBERTa, should be evaluated under the same protocol to test whether contextual embeddings provide stronger semantic generalization across languages.

Limitations and threats to validity

Several limitations affect the interpretation of the results. First, the Real class was formed from mainstream news portals and not from independently fact-checked texts. Therefore, the model may learn source and style differences rather than truth and falsity. Second, although the corpus is balanced, its size is limited: each language-class pair contains 452 texts, which may not cover the full diversity of topics. Third, structural differences between fact-checking websites and news portals may influence classification quality.

Fourth, the study uses only classical TF-IDF-based models. This is appropriate for the baseline objective, but it does not replace direct comparison with multilingual transformer models. Fifth, temporal and source-based splits were not tested separately; if training and test data are close in time or source structure, future performance may be lower. Sixth, exact duplicates were removed, but near-duplicates or event-level overlap may remain. Seventh, feature weights, important n-grams and SHAP-level explainability were not analyzed in this version.

Despite these limitations, the study offers an initial experimental basis for Kazakh-Russian fake/real classification. Its value is not in presenting a production detector, but in showing the level reached by an explainable baseline, revealing cross-lingual asymmetry and identifying the next steps for future research. The results can support corpus expansion, multi-level annotation and comparison with transformer-based models.

Conclusion. This paper evaluated TF-IDF-based machine learning models for fake/real text classification in Kazakh and Russian across five experimental scenarios. The balanced bilingual corpus contained 1,808 texts with equal language and class proportions. Logistic Regression, Linear SVM and Complement Naive Bayes were compared under a single evaluation protocol, and the results were interpreted as an explainable baseline.

In monolingual and bilingual regimes, the models achieved strong performance: Macro-F1 was approximately 0.985 in the bilingual, kk_only and ru_only scenarios. This indicates that TF-IDF-based linear models are a strong baseline for Kazakh-Russian fake news detection. At the same time, these scores do not mean that factual verification is solved, because the model may rely on style, topic and source signals. Cross-lingual evaluation showed a directional gap: Kazakh-to-Russian transfer reached Macro-F1 = 0.654, whereas Russian-to-Kazakh transfer reached Macro-F1 = 0.926.

Future work should expand the corpus, apply source-based and temporal splitting, and improve control over near-duplicate and event-level overlap. Multilingual transformer models such as XLM-RoBERTa and multilingual BERT should be compared with the TF-IDF baseline under the same protocol. The binary fake/real label should also be extended with multi-level annotation, including disinformation

technique, author intent, target audience and evidence, to improve the explanatory and applied value of the research.

Conflict of interests

The authors declare no conflict of interest.

Declaration on the use of generative AI tools

Generative AI tools were used only for language editing, readability improvement and technical formatting support under full author supervision. The tools were not used to generate the scientific content, methodology, dataset, experimental results, analysis or conclusions. All scientific content, references, data descriptions, results and conclusions were critically reviewed, verified and approved by the authors.

Acknowledgement and funding

The study was conducted without external funding. The authors have no additional acknowledgements.

References

Alghamdi J., Lin Y., & Luo S. (2026) Machine learning and deep learning approaches for fake news detection and related topics in multilingual contexts: A systematic literature review. Multimedia Tools and Applications, 85, Article 353. *https://doi.org/10.1007/s11042-026-21238-1* (in Eng.)

Conneau A., Khandelwal K., Goyal N., Chaudhary V., Wenzek G., Guzman F., Grave E., Ott M., Zettlemoyer L., & Stoyanov V. (2020) Unsupervised cross-lingual representation learning at scale. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. — P. 8440-8451. *https://doi.org/10.18653/v1/2020.acl-main.747* (in Eng.)

De, A., Bandyopadhyay D., Gain B., & Ekbal A. (2021) A transformer-based approach to multilingual fake news detection in low-resource languages. ACM Transactions on Asian and Low-Resource Language Information Processing, 21(1), Article 9. *https://doi.org/10.1145/3472619* (in Eng.)

Dementieva D., & Panchenko A. (2021) Cross-lingual evidence improves monolingual fake news detection. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing: Student Research Workshop. — P. 310-320. *https://doi.org/10.18653/v1/2021.acl-srw.32* (in Eng.)

Dementieva D., Kuimov M., & Panchenko A. (2023) Multiverse: Multilingual evidence for fake news detection. Journal of Imaging, 9(4), Article 77. *https://doi.org/10.3390/jimaging9040077* (in Eng.)

Devlin J., Chang M.-W., Lee K., & Toutanova K. (2019) BERT: Pre-training of deep bidirectional transformers for language understanding. Proceedings of NAACL-HLT 2019. — P. 4171-4186. *https://doi.org/10.18653/v1/N19-1423* (in Eng.)

Galli A., Masciari E., Moscato V., & Sperli G. (2022) A comprehensive benchmark for fake news detection. Journal of Intelligent Information Systems, 59. — P. 237-261. *https://doi.org/10.1007/s10844-021-00646-9* (in Eng.)

Gupta A., & Srikumar V. (2021) X-Fact: A new benchmark dataset for multilingual fact checking. Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 2: Short Papers). — P. 675-682. *https://doi.org/10.18653/v1/2021.acl-short.86* (in Eng.)

Hamed S.K., Ab Aziz M.J., & Yaakub M.R. (2023) A review of fake news detection approaches: A critical analysis of relevant studies and highlighting key challenges associated with the dataset, feature representation, and data fusion. Heliyon, 9(10), Article e20382. *https://doi.org/10.1016/j.heliyon.2023.e20382* (in Eng.)

Harris S., Hadi H.J., Ahmad N., & Alshara M.A. (2024) Fake news detection revisited: An extensive review of theoretical frameworks, dataset assessments, model constraints, and forward-looking

research agendas. Technologies, 12(11), Article 222. *https://doi.org/10.3390/technologies12110222* (in Eng.)

Providel E., Mendoza M., & Solar M. (2025) Cross-lingual cross-domain transfer learning for rumor detection. Future Internet, 17(7), Article 287. *https://doi.org/10.3390/fi17070287* (in Eng.)

Sambetbayeva M., Nekessova A., Yerimbetova A., Bayangali A., Kaldarova M., Telman D., & Smailov N. (2025) A multi-level annotation model for fake news detection: Implementing Kazakh-Russian corpus via Label Studio. Big Data and Cognitive Computing, 9(8), Article 215. *https://doi.org/10.3390/bdcc9080215* (in Eng.)

Shu K., Mahudeswaran D., Wang S., Lee D., & Liu H. (2020) FakeNewsNet: A data repository with news content, social context, and spatiotemporal information for studying fake news on social media. Big Data, 8(3). — P. 171-188. *https://doi.org/10.1089/big.2020.0062* (in Eng.)

Shu K., Sliva A., Wang S., Tang J., & Liu H. (2017) Fake news detection on social media: A data mining perspective. ACM SIGKDD Explorations Newsletter, 19(1). — P. 22-36. *https://doi.org/10.1145/3137597.3137600* (in Eng.)

Telman D., Yerimbetova A., Sambetbayeva M., & Bolatov B. (2026) Cross-lingual and multilingual approaches to fake news detection in the Kazakh language. Procedia Computer Science, 275. — P. 708-715. *https://doi.org/10.1016/j.procs.2026.01.082* (in Eng.)

Thibault C., Tian J.-J., Peloquin-Skulski G., Curtis T.L., Zhou J., Laflamme F., Guan Y., Rabbany R., Godbout J.-F., & Pelrine K. (2025) A guide to misinformation detection data and evaluation. Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining. — P. 5801-5809. *https://doi.org/10.1145/3711896.3737437* (in Eng.)

Thorne J., Vlachos A., Christodoulopoulos C., & Mittal A. (2018) FEVER: A large-scale dataset for fact extraction and verification. Proceedings of NAACL-HLT 2018. — P. 809-819. *https://doi.org/10.18653/v1/N18-1074* (in Eng.)

Vosoughi S., Roy D., & Aral S. (2018) The spread of true and false news online. Science, 359(6380). — P. 1146-1151. *https://doi.org/10.1126/science.aap9559* (in Eng.)

Wang W.Y. (2017) Liar, liar pants on fire: A new benchmark dataset for fake news detection. Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers). — P. 422-426. *https://doi.org/10.18653/v1/P17-2067* (in Eng.)

Zhou X., & Zafarani R. (2020) A survey of fake news: Fundamental theories, detection methods, and opportunities. ACM Computing Surveys, 53(5), Article 109. *https://doi.org/10.1145/3395046* (in Eng.)

Publication Ethics and Publication Malpractice
in the journals of the Central Asian Academic Research Center LLP

For information on Ethics in publishing and Ethical guidelines for journal publication see http://www.elsevier.com/publishingethics and http://www.elsevier.com/journal-authors/ethics.

Submission of an article to the journals of the Central Asian Academic Research Center LLP implies that the described work has not been published previously (except in the form of an abstract or as part of a published lecture or academic thesis or as an electronic preprint, see http://www.elsevier.com/postingpolicy), that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder. In particular, translations into English of papers already published in another language are not accepted.

No other forms of scientific misconduct are allowed, such as plagiarism, falsification, fraudulent data, incorrect interpretation of other works, incorrect citations, etc. The Central Asian Academic Research Center LLP follows the Code of Conduct of the Committee on Publication Ethics (COPE), and follows the COPE Flowcharts for Resolving Cases of Suspected Misconduct (http://publicationethics.org/files/u2/New_Code.pdf). To verify originality, your article may be checked by the Cross Check originality detection service http://www.elsevier.com/editors/plagdetect.

The authors are obliged to participate in peer review process and be ready to provide corrections, clarifications, retractions and apologies when needed. All authors of a paper should have significantly contributed to the research.

The reviewers should provide objective judgments and should point out relevant published works which are not yet cited. Reviewed articles should be treated confidentially. The reviewers will be chosen in such a way that there is no conflict of interests with respect to the research, the authors and/or the research funders.

The editors have complete responsibility and authority to reject or accept a paper, and they will only accept a paper when reasonably certain. They will preserve anonymity of reviewers and promote publication of corrections, clarifications, retractions and apologies when needed. The acceptance of a paper automatically implies the copyright transfer to the Central Asian Academic Research Center LLP.

The Editorial Board of the Central Asian Academic Research Center LLP will monitor and safeguard publishing ethics.

Requirements for articles design for publication in the journal are available on the websites:
www:nauka-nanrk.kz
http://physics-mathematics.kz/index.php/en/archive
ISSN2518-1726 (Online),
ISSN 1991-346X (Print)

Managing Editor: *A.Shormakova*
Editors: *D.S. Alenov, T. Apendiev*
Computer layout: *G.D. Zhadyranova*

Signed for print: June 15, 2026
Format: 70×90 1/16. 26.5 printed sheets. Order No. 2.

«Central Asian Academic Research Center» LLP
Almaty, Shevchenko Street, 28