Современная наука и инновации. 2025. № 1. С. 51-62.
ТЕХНИЧЕСКИЕ НАУКИ
ИНФОРМАТИКА, ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА И УПРАВЛЕНИЕ

Modern Science and Innovations. 2025;(1):51-62.
TECHNICAL SCIENCE
INFORMATICS, COMPUTER ENGINEERING AND MANAGEMENT

Научная статья
УДК 621.391
https://doi.org/10.37493/2307-910X.2025.1.4

Оценка бинарного прогнозирования мошеннических объявлений в облачных системах для отслеживания кандидатов ATS

Владимир Викторович Лиджи-Горяев[1], Галина Алексеевна Манкаева[2]*, Татьяна Борисовна Гольдварг[3], Светлана Сангаджиевна Мучкаева[4], Елена Николаевна Джахнаева[5]

[1, 2, 3, 4, 5] Калмыцкий государственный университет, г. Элиста, Россия
[1] vladlg@yandex.ru
[2] mankaeva.galina@yandex.ru
[3] tgoldvarg@bk.ru; https://orcid.org/0000-0002-4838-8783
[4] smuchkaeva@yandex.ru
[5] dzhakhnaeva-en@yandex.ru
*** Автор, ответственный за переписку:** Галина Алексеевна Манакаева, mankaeva.galina@yandex.ru

Аннотация. *Построение модели бинарной классификации для прогнозирования типа объявление в облачных ATS (системы отслеживания кандидатов) о приеме на работу (законные или мошеннические), может быть решено, в том числе, и с использованием различных алгоритмов машинного обучения. Для данной работы были выбраны традиционные алгоритмы классификации, включают LSVC (Linear Support Vector Classifier), GBT (Gradient Boost Tree) и RF(Random Forest). Один из подходов к построению такой модели заключается в выявлении и сборе соответствующих атрибутов или особенностей, которые могут помочь отличить мошенническое объявление о приеме на работу от законного. Некоторые из функций, которые могут быть полезны при обнаружении мошеннических объявлений о вакансиях, включают местоположение работы, описание вакансии, требования к работе, должностные обязанности, информацию о компании и данные о рекрутере. После чего различные алгоритмы машинного обучения могут быть обучены на подготовленных наборах данных с использованием стандартных методов, таких как перекрестная проверка, для оценки их производительности. Производительность обученных моделей может быть оценена с использованием различных оценочных показателей, таких как точность, прецизионность и отзывчивость. Наконец, наиболее эффективная модель может быть выбрана на основе оценочных показателей, а затем внедрена в производственную среду, где ее можно использовать для классификации объявлений о вакансиях как мошеннических или нет. Важно отметить, что модель также должна постоянно оцениваться и обновляться с течением времени, чтобы обеспечить ее надежность и эффективность. Исходя из результатов метрик оценки, был сделан вывод, что классификатор GBT показывает более высокую производительность и точность по сравнению с классификаторами LinearSVC и RF на данном наборе. Однако стоит учитывать, что классификатор GBT требует больше времени на обучение и прогнозирование, у GBT время 208.738579 с, а у LSVC и RF (64.267132 и 71.024914, соответственно). Учитывая результаты оценок для рабочей части программы использовали модель GBT. Для реализации прогнозирования было проведено машинное обучение на GBT, RF и LSVC на пользовательском наборе данных Job_Fraud, созданный на базе общедоступной EMSCAD. Для решения сильного дисбаланса данных была использована реализация на библиотеке синтетической передискретизации меньшинства (SMOTE). Сначала была получена модель, которую обучили на данных с помощью классификатора, удаление через TFIDFVectorizer стоп-слов в векторном пространстве, затем уменьшая разрядность данных перезагрузили данные, повторно*

обучили модель и векторайзер перед использованием их для прогнозирования. Для графического интерфейса использовался модуль tkinter. Функция predict() использует обученную модель для предсказаний на основе вектора признаков.

Ключевые слова: облачные ATS, детектирование мошеннических объявлений, классификаторы, модели LinearSVC, GBT и RF

Для цитирования: *Лиджи-Горяев В. В., Манкаева Г. А., Гольдварг Т. Б., Мучкаева С. С., Джахнаева Е. Н. Оценка бинарного прогнозирования мошеннических объявлений в облачных системах для отслеживания кандидатов ATS // Современная наука и инновации. 2025. № 1. С. 51-62. https://doi.org/10.37493/2307-910X.2025.1.4*

Research article

Assessment of binary prediction of fraudulent advertisements in ATS candidate tracking cloud systems

Vladimir V. Ligi-Goryaev[1], Galina A. Mankaeva[2]*, Tatyana B. Goldvarg[3], Svetlana S. Muchkaeva[4], Elena N. Dzhakhnaeva[5]

[1, 2, 3, 4, 5] Kalmyk State University, Elista, Russia
[1] vladlg@yandex.ru
[2] mankaeva.galina@yandex.ru
[3] tgoldvarg@bk.ru; https://orcid.org/0000-0002-4838-8783
[4] smuchkaeva@yandex.ru
[5] dzhakhnaeva-en@yandex.ru
*** Corresponding author:** Galina A. Mankaeva, mankaeva.galina@yandex.ru

Abstract. *The abstract describes the construction of a binary classification model for predicting the type of job advertisement in cloud-based ATS (Applicant Tracking Systems) as either legitimate or fraudulent. Various machine learning algorithms can be employed to address this issue. Traditional classification algorithms, including LSVC (Support Vector Machine), GBT (Gradient Boosting Tree), and RF (Random Forest), have been chosen for this study. One approach to building such a model involves identifying and collecting relevant attributes or features that can help distinguish fraudulent job advertisements from legitimate ones. Some features that could be useful in detecting fraudulent job ads include job location, job description, job requirements, job responsibilities, company information, and recruiter data. Subsequently, different machine learning algorithms can be trained on prepared datasets using standard methods such as cross-validation to assess their performance. The performance of the trained models can be evaluated using various metrics such as accuracy, precision, and recall. Ultimately, the most effective model can be selected based on these evaluation metrics and deployed in a production environment, where it can classify job advertisements as fraudulent or legitimate. It's important to note that the model should also undergo continuous evaluation and updates over time to ensure its reliability and effectiveness. Based on the evaluation metrics, it was concluded that the GBT classifier exhibits higher performance and accuracy compared to the LinearSVC and RF classifiers on the given dataset. However, it should be considered that the GBT classifier requires more time for training and prediction; GBT takes 208.738579 seconds, while LSVC and RF take 64.267132 and 71.024914 seconds, respectively. Taking into account the evaluation results, the GBT model was utilized for the operational aspect of the program. For implementation of the prediction, machine learning was performed on GBT, RF, and LSVC using a custom dataset called "Job_Fraud," created based on the publicly available EMSCAD dataset. To address the significant data imbalance, an implementation of the Synthetic Minority Over-sampling Technique (SMOTE) from a library was utilized. Initially, a model was obtained and trained on the data using a*

classifier, removing stop-words through TFIDFVectorizer in the vector space. Then, after reducing the dimensionality of the data, the data was reloaded, and both the model and vectorizer were retrained before being used for prediction. The tkinter module was used for the graphical interface. The predict() function utilizes the trained model for predictions based on the feature vector.

Keywords: cloud ATS, detection of fraudulent advertisements, classifiers, LinearSVC, GBT, RF models

For citation: *Ligi-Goryaev VM, Mankaeva GA, Goldvarg TB, Muchkaeva SS, Dzhakhnaeva EN. Assessment of binary prediction of fraudulent advertisements in ATS candidate tracking cloud systems. Modern Science and Innovations. 2025;(1):51-62. (In Russ.). https://doi.org/10.37493/2307-910X.2025.1.4*

Введение. Системы рекрутинга, также известные как системы отслеживания кандидатов (ATS), стали важным инструментом современного управления человеческими ресурсами. Эти системы оптимизируют процесс найма, помогая предприятиям эффективно управлять объявлениями о вакансиях, отслеживанием кандидатов, проверкой резюме и рабочими процессами найма. В связи с ростом популярности облачных решений целью данной статьи является изучение исследования рынка рекрутинговых систем (в том числе в России) с особым акцентом на функциональность облачных АТС и проблему мошенничества с вакансиями в автоматизированных системах подбора персонала. Ландшафт сервисов найма в России быстро развивается, и компании ищут инновационные и эффективные способы привлечения и удержания лучших специалистов. В последние годы набирает обороты внедрение рекрутинговых систем, в том числе облачных ATS, что позволяет пользователям получать к нему удаленный доступ через Интернет без необходимости локальной установки или обслуживания [1].

Функциональность облачных АТС в России. Функциональность облачных АТС в России значительно продвинулась вперед с несколькими ключевыми функциями, которые удовлетворяют уникальным требованиям российского рынка подбора персонала. Некоторые из примечательных функций облачной АТС в России включают в себя:

1. Настраиваемые рабочие процессы. Cloud ATS в России предлагает настраиваемые рабочие процессы, которые могут быть адаптированы в соответствии с конкретными потребностями и процессами найма российского бизнеса [2].
2. Локализация. Облачная АТС в России часто поставляется с функциями локализации, включая языковую поддержку, конвертацию валюты и соблюдение местных нормативных актов и законов о защите данных.

Разбор и скрининг резюме. Cloud ATS в России обычно включает в себя расширенные возможности анализа и проверки резюме, что позволяет эффективно управлять большими объемами резюме [3].

3. Интеграция с досками объявлений о вакансиях и соцсетями. Cloud ATS в России часто интегрируется с популярными досками объявлений о вакансиях и платформами социальных сетей, позволяя компаниям беспрепятственно публиковать вакансии, получать заявки и управлять взаимодействием с кандидатами. Это помогает расширить охват объявлений о вакансиях и, соответственно, привлечь более широкий круг кандидатов.

Рисунок 1 – Оценка развитости рекрутинговых сервисов по управлению кандидатами (ATS), используемых в России / Figure 1 – Assessment of the development of recruiting services for candidate management (ATS) used in Russia

Проблемы с мошенничеством с вакансиями в автоматизированных системах найма (ATS)

Несмотря на множество преимуществ использования автоматизированных систем найма, таких как ATS, существуют также проблемы, связанные с мошенничеством с вакансиями при трудоустройстве. Мошенничество с вакансиями относится к практике размещения поддельных вакансий или манипулирования процессом найма для личной выгоды [4]. Некоторые из проблем, связанных с мошенничеством при приеме на работу в автоматизированных системах найма (ATS), включают:

1. В некоторых случаях мошенники могут создавать поддельные объявления о вакансиях в автоматизированных системах найма, чтобы собирать личную информацию от ничего не подозревающих соискателей или выманивать у них деньги.
2. Автоматизированные системы подбора персонала полагаются на возможности анализа и проверки резюме для составления короткого списка кандидатов. Однако мошенники могут отправлять фальсифицированные резюме или манипулировать процессом проверки резюме, чтобы исказить свою квалификацию, навыки или опыт.
3. В автоматизированных системах найма могут отсутствовать эффективные механизмы проверки подлинности полномочий и квалификации кандидатов. Это может облегчить мошенникам подачу ложной информации или поддельных документов, что приведет к найму неквалифицированных или неподходящих кандидатов.
4. Автоматизированные системы подбора персонала хранят конфиденциальную личную и профессиональную информацию кандидатов, что делает их потенциальными целями для утечек данных и угроз безопасности. Мошенники могут использовать уязвимости в системе для получения несанкционированного доступа к персональным данным, что приводит к нарушениям конфиденциальности и репутационному ущербу для организации [5].

Решение проблемы мошенничества при приеме на работу в ATS. Организации должны проводить регулярные аудиты безопасности своих ATS для выявления и устранения потенциальных уязвимостей, которые могут быть использованы мошенниками. Это может включать в себя внедрение надежных мер безопасности, таких как шифрование, аутентификация и контроль доступа, а также системы фильтрации и прогнозирования мошеннических вакансий (СФМВ).

Современные мошенники умело используют те недостатки в области обеспечении безопасности в сфере применения современных средств связи, в том числе и сети Интернет, сегодня противодействия дистанционным мошенничествам (ДМ) развиваются и исследуются, они имеют возможности быстрой адаптации и динамики в зависимости от объекта внедрения. требует переосмысления тех социальных и правовых взаимоотношений которых присутствуют в цифровую эпоху. Одним из таких важных изменений от 29 ноября 2012 г. № 207-ФЗ «О внесении изменений в Уголовный кодекс Российской Федерации и отдельные законодательные акты Российской Федерации» [6], ответственность за противоправные действия законодатель разграничил по правоприменению в областях действия (ст. 159.1-159.6 УК РФ).

Материалы и методы исследований. В данной работе определяется оптимальный классификатор x, который максимизирует f(x)- целевую функцию, основанную на предварительно выбранных метриках. Задача f(x) может быть решена различными способами, в зависимости от конкретных требований и ограничений задачи. Например, можно определить f(x) как сумму точности и полноты классификации, или как долю правильно классифицированных мошеннических вакансий от общего числа мошеннических вакансий. При этом precision и recall часто используются совместно при оценке качества классификации, поскольку они предоставляют информацию о различных аспектах производительности модели [7].

В качестве целевой функции f(x) была выбрана формула:

$$F(x) = \alpha 1 \frac{TP}{FP + TP} + \alpha 2 \frac{TP}{TP + FN} \rightarrow max. \quad (1)$$

X – множество классификаторов, которые могут быть использованы для классификации вакансий

$\alpha 1$ и $\alpha 2$ – весовые коэффициенты с ограничением $\alpha 1 + \alpha 2 = 1$

В задаче обнаружения мошеннических вакансий в объявлениях, наиболее важным все же является метрика Recall, так как лучше пропустить несколько ложноположительных результатов (правильных вакансий, помеченных как мошеннические), чем пропустить реальную мошенническую вакансию (рис. 2).

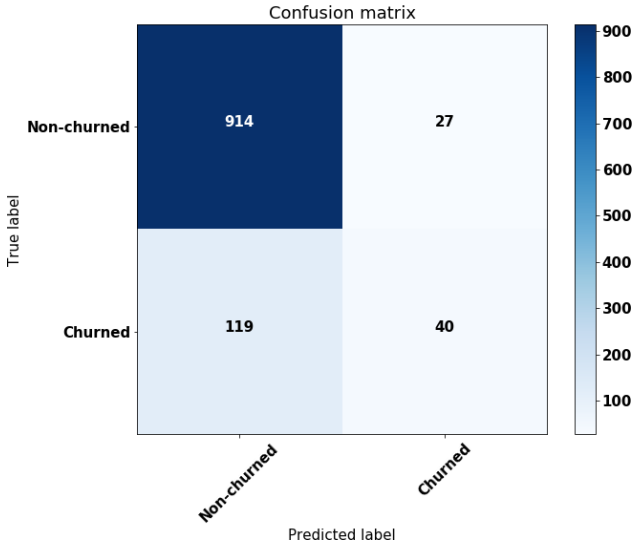

Рисунок 2 – Матрица путаницы / Figure 2 – Confusion Matrix

Для реализации прогнозирования было проведено машинное обучение на LR, RF и DT на EMSCAD [7]. Для решения сильного дисбаланса данных была использована реализация на библиотеке синтетической передискретизации меньшинства (SMOTE). Сначала была получена модель, которую обучили на данных с помощью классификатора, удаление через TFIDFVectorizer стоп-слов в векторном пространстве, затем уменьшая разрядность данных перезагрузили данные, повторно обучили модель и векторайзер перед использованием их для прогнозирования. Для графического интерфейса использовался модуль tkinter. Функция predict () использует обученную модель для прогноза на основе вектора признаков.

Пользовательский набор данных

Набор данных EMSCAD (Employer-Submitted Scam Corpus and Definitions) предоставляет ценный ресурс для обучения и оценки моделей машинного обучения для выявления мошеннических объявлений о вакансиях. Он состоит из 18 242 объявлений о приеме на работу, из которых 866 мошеннических и 17 376 законных. Набор данных легкодоступен и обеспечивает реалистичное представление о мошенничестве с трудоустройством в реальном мире, что делает его идеальным выбором для нашего исследования. Для улучшения локализации был создан пользовательский dataset Job_Fraud из 12 700 объявлений с 0,5% фейковых ваканский.

Извлечение признаков

Для разработки эффективной модели машинного обучения крайне важно извлечь значимые функции из набора данных Job_Fraud.

А.Предобработка

Вначале выполняется загрузка и подготовка данных. Код загружает набор данных о вакансиях из файла CSV, предварительно обрабатывает данные для создания нового столбца "текст" путем объединения нескольких других столбцов, а затем разбивает эти данные на обучающие и тестовые наборы. Текстовые данные в столбце "text" векторизуются с помощью CountVectorizer. Далее, модель обучается на основе обучающих данных и оценивается на основе тестовых данных.

```
df = pd.read_csv(job_fraud.csv')
# объединим несколько полей объявлений для каждой записи в одной строке
df['text'] = df[['title', 'location', 'department', 'company_profile', 'description', 'requirements', 'benefits']].apply(lambda x: ' '.join(str(i) for i in x if not pd.isnull(i)), axis=1)
```

Где все текстовые поля объединены в один столбец 'text'.

Создается функция, которая будет принимать текстовую строку и возвращать список признаков:

```
def get_features(text):
…
   return features
```

Эта функция извлекает средний вектор для всего документа и добавляет все географические сущности (GPE) как отдельные признаки, после чего можно будет использовать эту функцию для извлечения признаков из всех объявлений в искомом наборе данных:

```
y = df['fraudulent'].tolist()
for text in df['text'].tolist():
   X.append(get_features(text))
```

Б.Машинное обучение

Файл training.pkl и файл model.pkl - это два разных файла, которые имеют особенности своего назначения.

Training.pkl - это файл, который содержит данные, используемые для обучения модели машинного обучения. Он может включать в себя различные типы данных, такие как текстовые документы, изображения или числовые данные, которые были использованы для настройки параметров модели. Эти данные обычно используются для обучения модели на новых входных данных, что позволяет модели делать более точные прогнозы [9].

Model.pkl, с другой стороны, является файлом, который содержит обученную модель машинного обучения, которая была создана на основе данных, сохраненных в файле training.pkl. Эта модель может использоваться для предсказания результатов на новых данных, которые не были использованы при ее обучении. В отличие от training.pkl, model.pkl обычно уже не нуждается в изменении и может быть использован напрямую для получения результатов.

Таким образом, training.pkl и model.pkl - два важных компонента процесса машинного обучения, где первый используется для обучения модели, а второй - для применения обученной модели на новых данных.

Для получения модели model.pkl, нужно выполнить следующие шаги:

- Обучите модель машинного обучения на обучающих данных.
- Сохраните модель, используя функцию joblib.dump() и передать ей обученный объект модели и имя файла, в который нужно сохранить модель. Например:

joblib.dump (dt, 'model.pkl'), где «dt» в данном примере Decision Tree Classifier.

В. Дисбаланс данных

Чтобы решить проблему дисбаланса данных, можно использовать такие методы, как передискретизация или заниженная дискретизация. Избыточная выборка предполагает увеличение числа экземпляров в классе меньшинства, в то время как недостаточная выборка предполагает уменьшение числа экземпляров в классе большинствах [10].

Одним из популярных методов передискретизации является метод синтетической передискретизации меньшинства (SMOTE), который создает синтетические выборки путем интерполяции между соседними экземплярами класса меньшинства [11].

Для выполнения SMOTE библиотека imblearn на Python, которая предоставляет реализацию SMOTE:

```
from imblearn.over_sampling import SMOTE
smote = SMOTE(random_state=42)
X_train_resampled, y_train_resampled = smote.fit_resample(X_train_vec, y_train)
```

Этот код выполняет выборку обучающих данных с помощью SMOTE. Метод fit_resample принимает векторы объектов (X_train_vec) и метки (y_train) в качестве входных данных и возвращает пересмотренные версии обоих.

При решении проблемы дисбаланс часто сопутствующей проблемой может стать проблема разреженных данных (ПРД). Решить такую проблему можно решить, используя

методы уменьшения размерности, такие как анализ главных компонент (PCA) или усеченная декомпозиция по сингулярным значениям (SVD). Эти методы уменьшают размерность данных, сохраняя при этом большую часть дисперсии.

При этом надо преобразовать массив в список строк, прежде чем передавать его методу vectorizer fit, с помощью метода tolist() массива NumPy:

vectorizer.fit(X_train_transformed.tolist())

Результаты исследований и их обсуждение.

Таблица 1 – Метрики оценки классификаторов машинного обучения / Table 1 – Machine learning classifier evaluation metrics

	Классификаторы	**Метрики оценки**				
		Accuracy	**Precision**	**Recall**	**AUC**	**Время**
1	GBT (0.25%)	0,958420	0,924812	1,000000	0,997500	199,821130
2	GBT (0.50%)	0,951128	0,912916	0,997861	0,995594	208,738579
3	LinearSVC (0.50%)	0,923077	0,894339	0,963415	0,972366	64,267132
4	RF (0.50%)	0,921589	0,892537	0,959358	0,971597	71,024914

Из предоставленного вывода кода, можно сделать следующие выводы о производительности классификаторов GBT (Gradient Boosted Trees), LSVC(LinearSVC) и RF (Random Forest):

В обоих случаях для классификатора GBT (как 25% данных, так и 50% данных) показал более высокую точность (Accuracy) по сравнению с классификаторами LSVC и RF при использовании (табл.1).

Точность (Precision) классификации для класса мошенничества (метка 1.0) также была выше у классификаторов GBT по сравнению с другими классификаторами.

Показатель отзыва (Recall) классификации для класса мошенничества был высоким для всех классификаторов, близким к 1.0. Опять же, классификатор GBT имел немного более высокий показатель.

Площадь под кривой (AUC), которая является метрикой, измеряющей качество классификации, опять была высока для всех классификаторов. Однако классификатор GBT продемонстрировал немного более высокую AUC.

Время, затраченное на обучение и прогнозирование, было значительно больше для классификатора GBT по сравнению с классификаторами LinearSVC и RF. Исходя из этих результатов, можно сделать вывод, что классификатор GBT показывает более высокую производительность и точность по сравнению с классификаторами LinearSVC и RF на данном dataset. Однако стоит учитывать, что классификатор GBT требует больше времени на обучение и прогнозирование. Учитывая результаты оценок для рабочей части программы использовали модель GBT.

С. Предсказательная часть

В первой части кода были получены модели, которую только что обучили на данных с помощью классификатора, если загружать сохраненную модель и векторайзер без повторного обучения на новых данных, то в результате, если данные для предсказания имеют отличия от обучающих данных, то это обычно приводит к ошибкам классификации [12]. Для решения этой проблемы надо перезагрузить данные и повторно обучив модель и векторайзер перед использованием их для предсказания.

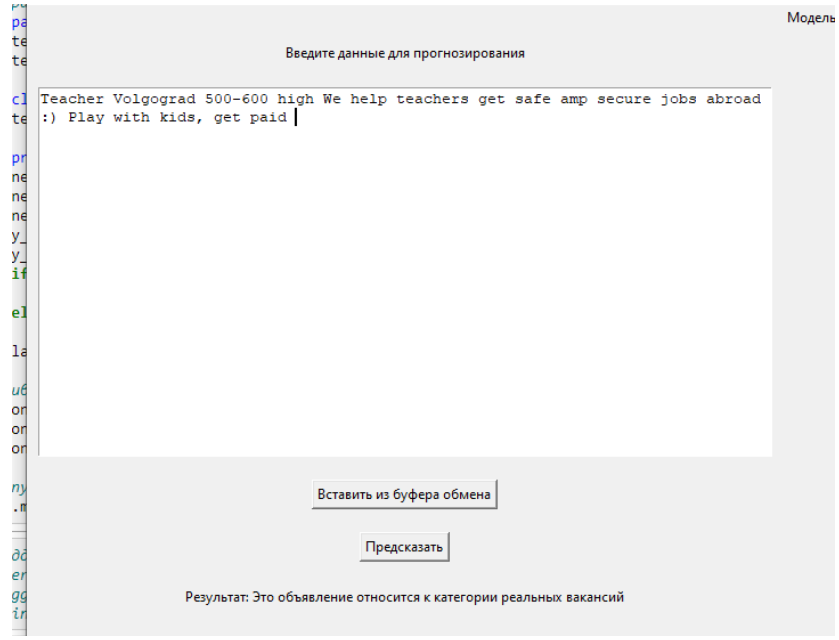

Рисунок 2 – Пример реальных объявлений / Figure 2 – Example of real ads

В коде импортируются необходимые библиотеки и создается графический интерфейс с помощью модуля tkinter. Функция predict() получает текст из виджета text_box, затем очищает его от знаков препинания и преобразует в вектор признаков с помощью сохраненного ранее объекта CountVectorizer (vect). Затем функция использует обученную модель, чтобы сделать предсказание на основе вектора признаков (рис. 2 и 3).

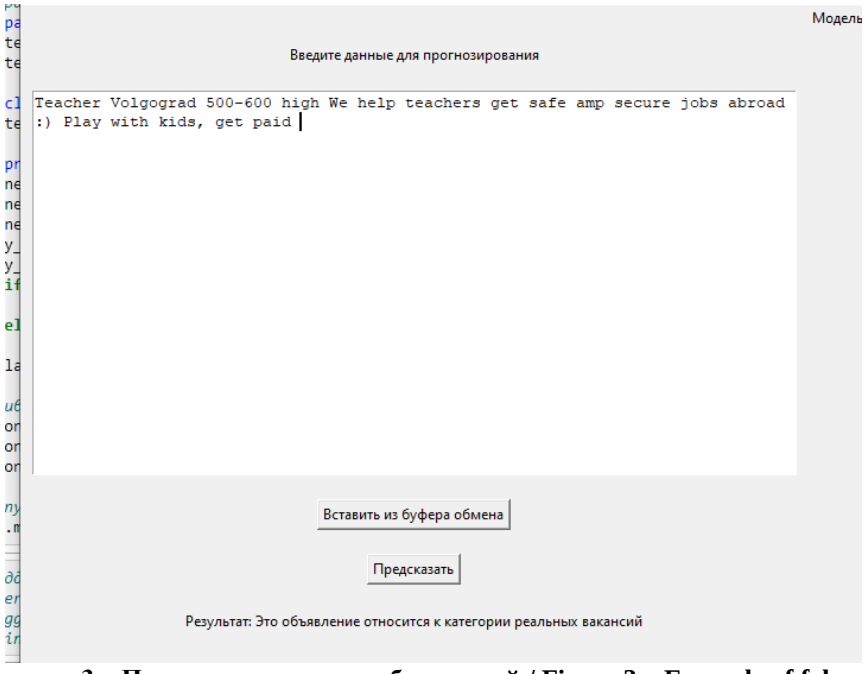

Рисунок 3 – Пример поддельных объявлений / Figure 3 – Example of fake ads

Заключение. Хотя автоматизированные системы найма, такие как ATS, предлагают многочисленные преимущества в оптимизации процесса найма, они также создают проблемы с точки зрения мошенничества при приеме на работу. Организациям необходимо знать о рисках, связанных с мошенничеством при приеме на работу и принимать упреждающие меры для снижения этих рисков.

В данной работе показатель классификатора GBT был лучше во всех компонентах метрик, которые использовались в программе. Однако время работы GBT 208.738579 в разы превышает время работы LSVC и RF (64.267132 и 71.024914, соответственно). Поэтому выбор между классификаторами может зависеть от баланса между точностью и временем выполнения в конкретной задаче.

Внедряя надежные процессы проверки, обеспечивая этичный и непредвзятое алгоритмическое проектирование, обучая менеджеров по найму, а также проводя регулярные аудиты безопасности, организации могут свести к минимуму влияние мошенничества и обеспечить справедливый и прозрачный процесс найма [14].

Проблемы при реализации кода возникли при обработке кириллице в собственном пользовательском наборе данных при обработке на платформе Apache Spark, однако, хотя задача вполне решаема и дальнейшая работа будет посвящена адаптации алгоритмов PySpark под русский алфавит.

Ссылка на код: https://github.com/GoryaevV/Job_Fraud.git

ЛИТЕРАТУРА

1. Настраиваемые рабочие процессы в облачных АТС в России. [Электронный ресурс]. URL: https://huntflow.ru/ (дата обращения 22.08.2023).
2. Исследование рынка рекрутинговых систем: функциональность облачных ATS в России. 02.11.2021. [Электронный ресурс]. URL: https://www.tadviser.ru/a/578060 (дата обращения 22.08.2023).
3. Скрининг-колл с рекрутером: вопросы, которые скорее всего вам зададут [Электронный ресурс]. URL: https://habr.com/ru/articles/689564//разборискрининг резюмевоблачныхатсвроссии (дата обращения 22.08.2023).
4. Swetha K., Sravani K. Fake job detection using machine learning approach // Journal of Engineering Sciences. 2023. Vol. 14. Issue 02. P. 67–74.
5. Бондарчук Д. В. Выбор оптимального метода интеллектуального анализа данных для подбора вакансий // Информационные технологии моделирования и управления. 2013. № 84 (6). С. 504–513.
6. Кудрявцев Р. В. Организация деятельности по раскрытию дистанционных мошенничеств // Молодой ученый. 2019. № 24 (262). С. 218–221. [Электронный ресурс]. URL: https://moluch.ru/archive/262/60528/ (дата обращения: 14.08.2023).
7. Горяев В. М., Бурлыков В. Д., Прошкин С. Н., Лиджи-Гаряев В. В., Джахнаева Е. Н. ROC-кривая и матрица путаницы как эффективное средство для оптимизации классификаторов машинного обучения // Вестник Башкирского университета. 2023. Т. 28. № 1. С. 22–28.
8. Лаборатория информационных и коммуникационных систем, Эгейский университет, Самос, Греция. Набор данных EMSCAD по мошенничеству при приеме на работу в Эгейском регионе. 2016. Доступно онлайн: http://icsdweb.aegean.gr/emscad (дата обращения 22.08.2023).
9. Горяев В. М., Басангова Е. О., Бембитов Д. Б., Мучкаева С. С., Сангаджиева С. В. Исследование производительности различных моделей машинного обучения при неинвазивном измерении артериального давления на основе сигналов PPG и ЭКГ // Вестник Башкирского университета. 2023. Т. 28. № 1. С. 36–44.
10. Wong Y., Kamel A. Classification of imbalanced data: a review. International Journal of Pattern Recognition and Artificial Intelligence. https://doi.org/10.1142/S0218001409007326
11. Tabassum H, Ghosh G. Detecting Online Recruitment Fraud Using Machine Learning, 2021 9th Int. Conf. Inf. Commun. Technol. ICoICT 2021. P. 472–477. https://doi.org/10.1109/ICoICT52021.2021.9527477
12. Борисов Е. С. Классификатор текстов на естественном языке. [Электронный ресурс]. URL: http://mechanoid.kiev.ua/neural-net-classifier-text.html

13. Коэльо Л. П., Ричарт В. Построение систем машинного обучения на языке Python. 2-е издание / Пер. с англ. Слинкин А. А. М.: ДМК Пресс, 2016. 302 с.
14. Горяев В. М. Разработка методики профессионально-психологического подбора кадров в организацию с учётом аспектов информационной безопасности // Современные наукоемкие технологии. 2021. № 12-2. С. 342–347.

REFERENCES

1. Customizable workflows in cloud PBX in Russia. Available from: https://huntflow.ru/ [Accessed 22 August 2023]. (In Russ.).
2. Research of the recruiting systems market: functionality of cloud ATS in Russia. 02.11.2021. Available from: https://www.tadviser.ru/a/578060 [Accessed 22 August 2023]. (In Russ.).
3. Screening call with a recruiter: questions that you are most likely to be asked. Available from: https://habr.com/ru/articles/689564//разборискрининг знакомствоблокныхатсвроссии [Accessed 22 August 2023]. (In Russ.).
4. Swetha K, Sravani K. Fake job detection using machine learning approach. Journal of Engineering Sciences. 2023;14(2):67-74.
5. Bondarchuk DV. Selecting the optimal method of data mining for job selection. Informatsionnye tekhnologii modelirovaniya i upravleniya = Modeling and Management of Information Technologies. 2013;84(6):504-513. (In Russ.).
6. Kudryavtsev RV. Organization of Activities to Detect Remote Frauds. Molodoi uchenyi = Young Scientist. 2019;24(262):218-221. Available from: https://moluch.ru/archive/262/60528/ [Accessed 14 August 2023]. (In Russ.).
7. Goryaev VM, Burlykov VD, Proshkin SN, Lidzhi-Garyaev VV, et al. ROC curve and confusion matrix as an effective tool for optimizing machine learning classifiers. Vestnik Bashkirskogo universiteta = Bulletin of Bashkir University. 2023;28(1):22-28. (In Russ.).
8. Laboratory of Information and Communication Systems, University of the Aegean, Samos, Greece. EMSCAD dataset on employment fraud in the Aegean region. 2016. Available from: http://icsdweb.aegean.gr/emscad [Accessed 22 August 2023]. (In Russ.).
9. Goryaev VM, Basangova EO, Bembitov DB, Muchkaeva SS, et al. Study of the performance of various machine learning models in non-invasive blood pressure measurement based on PPG and ECG signals. Bulletin of Bashkir University. 2023;28(1):36-44. (In Russ.).
10. Wong Y, Kamel A. Classification of imbalanced data: a review. International Journal of Pattern Recognition and Artificial Intelligence. https://doi.org/10.1142/S0218001409007326
11. Tabassum H, Ghosh G. Detecting Online Recruitment Fraud Using Machine Learning, 2021 9th Int. Conf. Inf. Commun. Technol. ICoICT 2021;472-477. https://doi.org/10.1109/ICoICT52021.2021.9527477
12. Borisov ES. Classifier of texts in natural language. Available from: http://mechanoid.kiev.ua/neural-net-classifier-text.html [Accessed 22 August 2023]. (In Russ.).
13. Coelho LP, Richart V. Building machine learning systems in Python. 2nd edition. Transl. from English. Slinkin AAM: DMK Press; 2016. 302 p. (In Russ.).
14. Goryaev VM. Development of a methodology for professional and psychological selection of personnel in an organization taking into account aspects of information security. Modern high technologies. 2021;(12-2):342-347. (In Russ.).

ИНФОРМАЦИЯ ОБ АВТОРАХ

Владимир Викторович Лиджи-Горяев – начальник отдела «Цифровая кафедра», Калмыцкий государственный университет имени Б.Б. Городовикова, +79371935125, vladlg@yandex.ru

Галина Алексеевна Манкаева – старший преподаватель, Калмыцкий государственный университет имени Б.Б. Городовикова, mankaeva.galina@yandex.ru

Татьяна Борисовна Гольдварг – доцент кафедры экспериментальной физики, Калмыцкий государственный университет имени Б.Б. Городовикова, +79093974451, tgoldvarg@bk.ru

Светлана Сангаджиевна Мучкаева – доцент кафедры алгебры и анализа, Калмыцкий государственный университет имени Б.Б. Городовикова, +79054007024, smuchkaeva@yandex.ru

Елена Николаевна Джахнаева – старший преподаватель, Калмыцкий государственный университет имени Б.Б. Городовикова, +79371927755, dzhakhnaeva-en@yandex.ru

Вклад авторов: все авторы внесли равный вклад в подготовку публикации.
Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Статья поступила в редакцию: 14.02.2025;
одобрена после рецензирования: 15.04.2025;
принята к публикации: 22.04.2025.

INFORMATION ABOUT THE AUTHORS

Vladimir V. Ligi-Goryaev – Head of the Digital Department, Kalmyk State University named after B.B. Gorodovikov, +79371935125, vladlg@yandex.ru

Galina A. Mankaeva – Senior Lecturer, Kalmyk State University named after B.B.Gorodovikov, +79061764200, mankaeva.galina@yandex.ru

Tatyana B. Goldvarg – Associate Professor of the Department of Experimental Physics, Kalmyk State University named after B.B. Gorodovikov, +79093974451, tgoldvarg@bk.ru

Svetlana S. Muchkaeva – Associate Professor of the Department of Algebra and Analysis, Kalmyk State University named after B.B. Gorodovikov, +79054007024, smuchkaeva@yandex.ru

Elena N. Dzhakhnaeva – Senior Lecturer, Kalmyk State University named after B.B. Gorodovikov, +79371927755, dzhakhnaeva-en@yandex.ru

Contribution of the authors: the authors contributed equally to this article.
Conflict of interest: the authors declare no conflicts of interests.

The article was submitted: 14.02.2025;
approved after reviewing: 15.04.2025;
accepted for publication: 22.04.2025.