

САНКТ-ПЕТЕРБУРГСКИЙ ПОЛИТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ
ПЕТРА ВЕЛИКОГО

Физико-механический институт

Высшая школа прикладной математики и вычислительной физики

ОТЧЁТ ПО ЛАБОРАТОРНОЙ РАБОТЕ № 2

«Перевод аннотаций»

по дисциплине «Автоматизация научных исследований»

Выполнил:

Смирнов Александр Дмитриевич

студент гр. 5040102/50201

Преподаватель: Новиков Ф.А.

Санкт-Петербург

2025 г.

1. Задание

В рамках лабораторной работы необходимо:

1. Найти любые две аннотации к научным статьям с достаточным количеством научных терминов, написанные на русском языке.
2. Перевести выбранные аннотации на английский язык с помощью ИИ, используя предложенные промпты (5 промптов). Использовать не менее двух ИИ-моделей.
3. Для пятого промпта выполнить оба этапа. В третьем и пятом промптах сделать выбор среди выделенного жирным (контекст и стиль; слова и значения).
4. При необходимости использовать уточняющие промпты и указать их в выводах.
5. Провести сравнительный анализ полученных результатов по критериям качества.
6. Написать собственный промпт (Промпт 6) для создания аналогичной аннотации на английском языке с сохранением смысла.
7. В выводах указать рекомендуемый вариант, цепочку промптов и обоснование выбора.

2. Используемые модели ИИ

В работе использованы две ИИ-модели в рамках платформы ChatGPT:

- Модель А: ChatGPT (GPT-5.2 Thinking)
- Модель В: ChatGPT (GPT-5.2, быстрый режим генерации)

3. Промпты

Промпт 1 (Базовый):

Переведи этот текст на английский: {текст}

Промпт 2 (С фокусом на качество):

Переведи следующий текст с русского на английский. Следи за тем, чтобы грамматика была правильной, а перевод звучал естественно для носителя языка. Вот текст: {текст}

Промпт 3 (Комплексный и структурированный):

Ты — профессиональный переводчик... Учти:

- 1) Контекст: Текст относится к [указать область].

- 2) Терминология: Используй точные профессиональные термины.
 - 3) Стилъ и тон: Используй [официальный / неформальный / научный / художественный] стилъ.
 - 4) Грамматика и идиомы: избегай дословного перевода.
- При необходимости — сноски.

Промпт 4 (Анализирующий и автономный):

Проанализируй область, стилъ, терминологию и ключевые слова, затем выполни перевод; при необходимости — сноски.

Промпт 5 (Двухэтапный):

Этап 1 — выдели неоднозначные слова (не фразы) и оформи таблицей (слово / варианты перевода / рекомендованный).

Этап 2 — выполни перевод с учетом выбранных значений для перечисленных слов.

Промпт 6 (Индивидуальный):

Ты — научный редактор англоязычных статей. Перепиши следующую русскоязычную аннотацию на английском языке как оригинальную (не дословный перевод), сохранив смысл, структуру и научный стилъ.

Требования: (1) 120–180 слов; (2) корректная терминология; (3) логичная связность; (4) в конце добавь строку Keywords: ... Русский текст:

{текст}

4. Исходные аннотации и источники

Аннотация 1 (источник: CEUR Workshop Proceedings, Vol-1864, paper_33.pdf).

Аннотация— Данная статья посвящена исследованию различных подходов к идентификации аномалий во временных рядах, которая заключается в обнаружении и обработке отклонений в потоках данных, получаемых во время проведения технологических процессов.

Выявление аномалий в поведении системы позволяет не только повысить качество таких процессов, но и предотвращать нештатные ситуации и аварии на ранних этапах. Все это указывает на актуальность

проведения исследований в данной области. В работе приведен обзор существующих методов и алгоритмов обнаружения аномалий с целью структуризации имеющихся данных и последующего отбора средств для разработки системы идентификации аномалий в потоках больших данных.

Ключевые слова — Поиск аномалий, Анализ данных, Потоки данных.

Аннотация 2 (источник: журнал «Прикладная статистика и искусственный интеллект», 2025 №2, 6-2025-chernokozhih.pdf).

Аннотация. В исследовании построены модели бинарной классификации для прогнозирования оттока клиентов в телекоммуникационной компании методами машинного обучения при помощи языка программирования Python. В качестве факторов были взяты личные данные клиента, финансовые характеристики, информация о подключениях и услугах. Датасет был проанализирован на наличие дубликатов, пустых значений, было осуществлено преобразование некоторых полей в более удобный для моделирования тип данных, введены фиктивные переменные и устранен дисбаланс выборки методом RandomOverSampler. Были разработаны и обучены несколько моделей машинного обучения, включая логистическую регрессию, решающие деревья, модель опорных векторов и градиентные бустинги. Была оценена их производительность по различным метрикам и был выбран в качестве наиболее эффективной модели градиентный бустинг для прогнозирования оттока клиентов. Наибольший вклад в прогноз внесли следующие признаки: общие выплаты, размер последней месячной выплаты, частота оплаты раз в 2 года и наличие у абонента услуги домашнего интернета. Практическая значимость данного исследования заключается в разработке стратегии по удержанию новых клиентов и совершенствованию сервиса телекоммуникационных компаний. Это подразумевает улучшение информированности абонентов о предоставляемых услугах и подбор оптимальных тарифных планов, исходя из их потребностей и финансовых возможностей.

Ключевые слова: классификация, телекоммуникации, прогнозирование

оттока клиентов, методы машинного обучения, абоненты, градиентный бустинг.

5. Результаты генерации переводов

5.1. Аннотация 1

Промпт 1. Базовый

Промпт:

Переведи этот текст на английский: {Аннотация 1}

Модель A: ChatGPT (GPT-5.2 Thinking) — Ответ:

Abstract—This paper investigates various approaches to anomaly identification in time series, which involves detecting and processing deviations in data streams obtained during technological processes. Detecting anomalies in system behavior not only improves the quality of such processes, but also helps prevent abnormal situations and accidents at early stages. This underscores the relevance of research in this area. The paper reviews existing anomaly detection methods and algorithms in order to structure the available knowledge and to select tools for developing an anomaly identification system for big-data streams.

Keywords—anomaly detection, data analysis, data streams.

Модель B: ChatGPT (GPT-5.2, быстрый режим генерации) — Ответ:

Abstract—The article studies different approaches to identifying anomalies in time series, i.e., the detection and handling of deviations in data streams generated by technological processes. Anomaly detection in system behavior can improve process quality and enable early prevention of emergency situations and accidents, which makes this research area highly relevant. The paper provides an overview of existing anomaly-detection methods and algorithms to organize current knowledge and support the selection of techniques for building anomaly identification systems for big data streams.

Keywords—anomaly search, data analysis, data streams.

Промпт 2. С фокусом на качество

Промпт:

Переведи следующий текст с русского на английский... Вот текст:
{Аннотация 1}

Модель А: ChatGPT (GPT-5.2 Thinking) — Ответ:

This paper explores approaches to anomaly identification in time series, focusing on detecting and processing deviations in data streams recorded during technological processes. Recognizing anomalies in system behavior can both improve process quality and help prevent abnormal events and accidents at early stages, highlighting the importance of research in this area. The study surveys existing anomaly detection methods and algorithms, structures the accumulated knowledge, and outlines how these results can support the selection of tools for developing anomaly identification systems for big-data streams.

Модель В: ChatGPT (GPT-5.2, быстрый режим генерации) — Ответ:

The paper examines approaches to identifying anomalies in time series by detecting and processing deviations in data streams collected during technological processes. Detecting anomalies in system behavior can enhance process quality and help avert emergencies and accidents early, which confirms the relevance of the topic. The work reviews current anomaly-detection methods and algorithms, organizes existing findings, and provides a basis for choosing techniques to design anomaly identification systems for large-scale data streams.

Промпт 3. Комплексный и структурированный

Промпт:

Контекст: область интеллектуального анализа данных и анализа временных рядов (anomaly detection).

Стиль и тон: научный.

Переведи текст: {Аннотация 1}

Модель A: ChatGPT (GPT-5.2 Thinking) — Ответ:

Abstract—This paper addresses time-series anomaly detection in the context of data mining for industrial and technological process monitoring. Accurate identification and treatment of deviations in streaming data can improve process performance and enable early prevention of abnormal events and accidents. The study reviews and systematizes existing anomaly detection methods and algorithms to structure the available body of knowledge and to support the selection of techniques for developing anomaly identification systems for big-data streams.

Keywords—anomaly detection; time series; data mining; data streams.

Модель B: ChatGPT (GPT-5.2, быстрый режим генерации) — Ответ:

Abstract—The text relates to data mining and time-series analytics. It focuses on detecting and handling deviations in streaming data produced by technological processes. Timely anomaly detection helps improve process quality and prevents abnormal situations and accidents at early stages. The paper surveys and organizes existing anomaly-detection methods and algorithms to support the choice of tools for building anomaly identification systems for big-data streams.

Keywords—time series; anomaly detection; data streams; data mining.

Промпт 4. Анализирующий и автономный

Промпт:

Проанализируй предоставленный текст и выполни перевод:

{Аннотация 1}

Анализ:

- 1) Тематическая область: интеллектуальный анализ данных (data mining), анализ временных рядов, обнаружение аномалий в потоках данных.
- 2) Стил ь и тон: научный, информативный, без эмоциональной окраски.
- 3) Ключевая терминология: аномалии, временные ряды, потоки данных, технологические процессы, обнаружение/идентификация,

методы и алгоритмы.

Потенциально сложные элементы: термины

«идентификация/обнаружение», «потoki больших данных».

4) Ключевые слова: anomaly detection, time series, data streams, data mining, technological processes.

Модель А: ChatGPT (GPT-5.2 Thinking) — Ответ:

Abstract—The paper investigates approaches to anomaly identification in time series by detecting and processing deviations in data streams produced during technological processes. Detecting anomalies in system behavior can improve process quality and help prevent abnormal situations and accidents at early stages, which underlines the relevance of research in this field. The study provides a survey of existing anomaly detection methods and algorithms, structures the accumulated knowledge, and supports the selection of tools for developing anomaly identification systems for big-data streams.

Keywords—anomaly detection; time series; data streams; data mining.

Модель В: ChatGPT (GPT-5.2, быстрый режим генерации) — Ответ:

Abstract—This work analyzes approaches to anomaly identification in time series, i.e., the detection and processing of deviations in data streams generated in technological processes. Anomalies in system behavior, once detected, can improve process quality and enable early prevention of emergencies and accidents. The paper reviews existing anomaly-detection methods and algorithms, systematizes current results, and assists in selecting tools for developing anomaly identification systems for high-volume data streams.

Keywords—anomaly detection; time series; data streams.

Промпт 5. Двухэтапный

Промпт:

Этап 1: выдели неоднозначные слова. Этап 2: выполни перевод с учетом выбранных значений.

Текст: {Аннотация 1}

Этап 1. Таблица неоднозначных слов:

Слово (RU)	Варианты перевода (EN) + контекст	Рекомендуемый вариант
подходов	approaches / methods (в научном контексте: approaches)	approaches
идентификации	identification / detection (в задачах аномалий чаще: detection)	detection
обработке	processing / handling / treatment	processing
отклонений	deviations / abnormalities / outliers	deviations
потоках	streams / flows	streams
средств	tools / means / instruments	tools
разработки	development / design	development

Этап 2. Выбранные значения слов для перевода:

- ["идентификации"] → detection
- ["отклонений"] → deviations
- ["потоках"] → streams
- ["средств"] → tools

Модель A: ChatGPT (GPT-5.2 Thinking) — Ответ:

Abstract—This paper investigates approaches to anomaly detection in time series, focusing on detecting and processing deviations in data streams obtained during technological processes. Detecting anomalies in system

behavior can improve process quality and help prevent abnormal situations and accidents at early stages, which highlights the relevance of this research area. The paper reviews existing anomaly detection methods and algorithms to structure the current knowledge base and to support the selection of tools for developing an anomaly identification system for big-data streams.

Keywords—anomaly detection; time series; data streams.

Модель В: ChatGPT (GPT-5.2, быстрый режим генерации) —

Ответ:

Abstract—The article is devoted to approaches to anomaly detection in time series and considers the detection and processing of deviations in data streams that arise during technological processes. Identifying anomalies in system behavior can increase the quality of such processes and allow abnormal situations and accidents to be prevented at early stages. The paper provides an overview of existing methods and algorithms to systematize available information and to choose tools for developing anomaly identification systems for large data streams.

Keywords—anomaly detection; data streams; time series.

Промпт 6. Индивидуальный

Промпт:

Ты — научный редактор англоязычных статей. Перепиши следующую русскоязычную аннотацию на английском языке как оригинальную (не дословный перевод), сохранив смысл, структуру и научный стиль.

Требования: (1) 120–180 слов; (2) корректная терминология; (3) логичная связность; (4) в конце добавь строку Keywords: ... Русский текст:

{Аннотация 1}

Модель А: ChatGPT (GPT-5.2 Thinking) — Ответ:

Anomaly detection in time series is essential for monitoring technological processes, where data are generated as continuous streams. Deviations in such streams may indicate measurement errors, outliers, or early signs of abnormal system behavior. Timely identification and processing of

anomalies can improve the reliability of analytical results, enhance process performance, and reduce the risk of emergencies and accidents. This paper reviews and systematizes existing anomaly detection methods and algorithms with the aim of structuring the current state of the field. The resulting overview supports the selection of practical techniques for designing anomaly identification systems intended to operate on high-volume, streaming data.

Keywords: anomaly detection; time series; data streams; data mining.

5.2. Аннотация 2

Промпт 1. Базовый

Промпт:

Переведи этот текст на английский: {Аннотация 2}

Модель А: ChatGPT (GPT-5.2 Thinking) — Ответ:

Abstract. The study develops binary classification models to predict customer churn in a telecommunications company using machine-learning methods implemented in Python. Customer personal data, financial characteristics, and information on connections and subscribed services are used as predictors. The dataset is checked for duplicates and missing values; several fields are transformed into types more suitable for modelling, dummy variables are introduced, and class imbalance is addressed with RandomOverSampler. Multiple machine-learning models are trained, including logistic regression, decision trees, support vector machines, and gradient boosting. Performance is evaluated using several metrics, and gradient boosting is selected as the most effective model for churn prediction. The strongest contributors to the prediction are total charges, the latest monthly charge, the frequency of payment once every two years, and whether the subscriber has home internet service. The practical value of the study lies in developing retention strategies for new customers and improving telecom service quality, including better customer awareness and selecting optimal tariff plans based on needs and financial capacity.

Keywords: classification; telecommunications; customer churn prediction; machine learning; subscribers; gradient boosting.

Модель B: ChatGPT (GPT-5.2, быстрый режим генерации) — Ответ:

Abstract. This research builds binary classification models for forecasting customer churn in a telecom company using machine learning in Python. Predictors include customers' personal information, financial indicators, and data on connections and services. The dataset is preprocessed by detecting duplicates and missing values, converting some fields to modelling-friendly formats, creating dummy variables, and handling class imbalance with RandomOverSampler. Several models are trained—logistic regression, decision trees, support vector machines, and gradient boosting methods. After evaluation with multiple quality metrics, gradient boosting is chosen as the most effective approach. Key features are total payments, the last monthly payment amount, a two-year payment frequency, and the presence of home internet service. The results can be used to design customer retention strategies and improve telecom service by informing subscribers and selecting suitable tariff plans.

Keywords: classification; telecom; churn; machine learning; gradient boosting.

Промпт 2. С фокусом на качество

Промпт:

Переведи следующий текст с русского на английский... Вот текст:
{Аннотация 2}

Модель A: ChatGPT (GPT-5.2 Thinking) — Ответ:

This study builds binary classification models to predict customer churn in a telecommunications company using machine learning implemented in Python. Predictors include customers' personal data, financial characteristics, and information about connections and subscribed services. The dataset is cleaned (duplicates and missing values), some variables are converted to modelling-friendly data types, dummy variables are created, and class imbalance is corrected via RandomOverSampler. Several models are trained—logistic regression, decision trees, support vector machines, and gradient boosting—and compared across multiple performance metrics.

Gradient boosting demonstrates the best overall results and is selected as the final model. The most influential features include total charges, the latest monthly charge, a two-year payment frequency, and the presence of home internet service. The findings support the development of retention strategies and service improvements in telecom companies.

Модель B: ChatGPT (GPT-5.2, быстрый режим генерации) —

Ответ:

The paper translates the churn prediction task into binary classification and trains models for a telecommunications company using Python-based machine learning. It uses customer demographics, financial variables, and connection/service attributes. Data preparation includes removing duplicates, handling missing values, transforming several fields, creating dummy variables, and balancing classes with RandomOverSampler. The study compares logistic regression, decision trees, SVM, and gradient boosting models using several metrics. Gradient boosting performs best and is selected for churn prediction. The most important predictors are total payments, the last monthly payment, payment frequency once every two years, and having home internet. The results can inform customer retention programs and help tailor tariff plans to customers' needs and budgets.

Промпт 3. Комплексный и структурированный

Промпт:

Контекст: прикладное машинное обучение в телекоммуникациях (customer churn prediction).

Стиль и тон: научный.

Переведи текст: {Аннотация 2}

Модель A: ChatGPT (GPT-5.2 Thinking) — Ответ:

Abstract—The text relates to applied machine learning for telecommunications analytics (customer churn prediction). Use accurate terms common in predictive modelling and evaluation.

The study builds binary classification models to predict customer churn in a

telecommunications company using Python-based machine learning. Predictors include customer demographics, financial variables, and information on connections and services. The dataset is preprocessed by removing duplicates and missing values, converting selected fields into modelling-friendly formats, generating dummy variables, and addressing class imbalance with RandomOverSampler. Multiple models are trained—logistic regression, decision trees, support vector machines, and gradient boosting—and evaluated using several metrics. Gradient boosting is selected as the most effective approach. The strongest contributors to the prediction are total charges, the latest monthly charge, a two-year payment frequency, and the availability of home internet service.

Keywords—customer churn; binary classification; telecommunications; machine learning; gradient boosting.

Модель В: ChatGPT (GPT-5.2, быстрый режим генерации) —

Ответ:

Abstract—Domain: telecom analytics and machine learning. Style: scientific.

This study develops binary classification models for forecasting customer churn in a telecommunications company using machine learning in Python. Features include customer personal data, financial characteristics, and information about connections and subscribed services. Data preprocessing covers duplicate and missing-value handling, type conversion, dummy-variable encoding, and class rebalancing via RandomOverSampler. The authors train logistic regression, decision tree, SVM, and gradient boosting models and compare them using multiple performance metrics. Gradient boosting is chosen as the most effective model. The most influential predictors are total payments, last monthly payment amount, a two-year payment schedule, and home internet service availability.

Keywords—churn prediction; telecom; machine learning; gradient boosting.

Промпт 4. Анализирующий и автономный

Промпт:

Проанализируй предоставленный текст и выполни перевод:
{Аннотация 2}

Анализ:

- 1) Тематическая область: прикладное машинное обучение, предиктивная аналитика в телекоммуникациях (прогнозирование оттока).
- 2) Стилъ и тон: научный, отчетный.
- 3) Ключевая терминология: бинарная классификация, предобработка данных, фиктивные переменные (dummy variables), дисбаланс классов, RandomOverSampler, логистическая регрессия, решающие деревья, метод опорных векторов (SVM), градиентный бустинг, метрики качества, важность признаков.
Потенциально сложные элементы: корректный перевод «отток клиентов» (customer churn), «фиктивные переменные» (dummy variables), названия алгоритмов.
- 4) Ключевые слова: churn prediction, binary classification, data preprocessing, class imbalance, gradient boosting, telecom analytics.

Модель А: ChatGPT (GPT-5.2 Thinking) — Ответ:

Abstract. The study develops binary classification models for predicting customer churn in a telecommunications company using machine learning in Python. Predictors include customer demographics, financial characteristics, and information on connections and services. Data preprocessing involves checking for duplicates and missing values, converting selected variables to modelling-friendly types, creating dummy variables, and mitigating class imbalance using RandomOverSampler. Several models are trained and compared, including logistic regression, decision trees, support vector machines, and gradient boosting. Based on multiple evaluation metrics, gradient boosting is selected as the most effective model for churn prediction. The most influential features are total charges, the latest monthly charge, a two-year payment frequency, and the presence of home internet service. The results support the design of customer retention strategies and improvements in telecom services.

Keywords: customer churn; binary classification; machine learning; telecommunications; gradient boosting.

**Модель В: ChatGPT (GPT-5.2, быстрый режим генерации) —
Ответ:**

Abstract. This work applies machine learning to customer churn prediction in a telecommunications company and formulates the task as binary classification implemented in Python. The feature set includes personal customer attributes, financial indicators, and connection/service information. The dataset is cleaned, variables are transformed, dummy encoding is performed, and class imbalance is addressed with RandomOverSampler. Logistic regression, decision trees, SVM, and gradient boosting models are trained and evaluated with multiple metrics. Gradient boosting is chosen as the best-performing model. The largest contribution to the prediction comes from total payments, last monthly payment, a two-year billing cycle, and the availability of home internet. The findings can be used to develop retention programs and improve service offerings.

Keywords: churn prediction; telecom analytics; machine learning; gradient boosting.

**Промпт 5. Двухэтапный
Промпт:**

Этап 1: выдели неоднозначные слова. Этап 2: выполни перевод с учетом выбранных значений.

Текст: {Аннотация 2}

Этап 1. Таблица неоднозначных слов:

Слово (RU)	Варианты перевода (EN) + контекст	Рекомендуемый вариант
факторы	factors / predictors / features	predictors
услугах	services / utilities	services
преобразование	transformation / conversion	conversion

фиктивные	dummy / fictitious	dummy
дисбаланс	imbalance / skew	imbalance
производительность	performance / efficiency	performance
вклад	contribution / impact	contribution
подбор	selection / matching / tuning	selection

Этап 2. Выбранные значения слов для перевода:

- ["факторы"] → predictors
- ["фиктивные"] → dummy
- ["производительность"] → performance
- ["вклад"] → contribution

Модель A: ChatGPT (GPT-5.2 Thinking) — Ответ:

Abstract. The study builds binary classification models to predict customer churn in a telecommunications company using Python-based machine learning. Customer personal data, financial characteristics, and information on connections and subscribed services are used as predictors. The dataset is cleaned (duplicates and missing values), selected fields are converted to modelling-friendly types, dummy variables are introduced, and class imbalance is corrected with RandomOverSampler. Several models are trained and evaluated, including logistic regression, decision trees, support vector machines, and gradient boosting. Based on multiple metrics, gradient boosting is selected as the most effective model. The largest contribution to the prediction comes from total charges, the latest monthly charge, a two-year payment frequency, and the presence of home internet service. Practical significance includes developing retention strategies and improving telecom services through better customer awareness and tariff plan selection. Keywords: customer churn; binary classification; data preprocessing; gradient boosting.

**Модель В: ChatGPT (GPT-5.2, быстрый режим генерации) —
Ответ:**

Abstract. This research constructs binary classification models for customer churn prediction in a telecom company using machine learning in Python. Predictors include personal customer attributes, financial indicators, and data on connections and services. The dataset is preprocessed by removing duplicates and missing values, converting several fields, applying dummy-variable encoding, and eliminating class imbalance with RandomOverSampler. Multiple models—logistic regression, decision trees, SVM, and gradient boosting—are trained and assessed across several metrics. Gradient boosting is chosen as the best-performing approach. The most important predictors are total payments, last monthly payment, a two-year billing frequency, and home internet service. The results inform customer retention programs and tariff plan selection.

Keywords: churn prediction; machine learning; telecommunications; gradient boosting.

**Промпт 6. Индивидуальный
Промпт:**

Ты — научный редактор англоязычных статей. Перепиши следующую русскоязычную аннотацию на английском языке как оригинальную (не дословный перевод), сохранив смысл, структуру и научный стиль.

Требования: (1) 120–180 слов; (2) корректная терминология; (3) логичная связность; (4) в конце добавь строку Keywords: ... Русский текст:

{Аннотация 2}

Модель А: ChatGPT (GPT-5.2 Thinking) — Ответ:

Customer churn is a major challenge in telecommunications, where competitive pressure makes retention strategies crucial. This study frames churn prediction as a binary classification task and builds models in Python using customer demographics, financial variables, and connection/service attributes. The data are preprocessed by handling duplicates and missing values, converting selected fields, applying dummy encoding, and correcting

class imbalance with RandomOverSampler. Several algorithms are trained and compared, including logistic regression, decision trees, support vector machines, and gradient boosting. Using multiple evaluation metrics, gradient boosting is selected as the best-performing model. Feature importance analysis highlights total charges, the latest monthly charge, a two-year payment schedule, and home internet service as key predictors. The findings support targeted retention initiatives and better tariff plan recommendations. Keywords: customer churn; telecom analytics; binary classification; data preprocessing; gradient boosting.

6. Сравнительный анализ

Оценивание выполнено по критериям (0–10): грамматика, точность терминологии, естественность, сохранение стиля, учет контекста, работа с идиомами.

Поскольку тексты научные и практически не содержат идиом, критерий 6 в основном отражает отсутствие дословности и корректность адаптации устойчивых выражений.

6.1. Оценка переводов Аннотации 1

Вариант	Грамм.	Термин.	Естеств.	Стиль	Смысл	Идиомы	Итого/60
П1 — А	9	8	8	8	8	10	51
П1 — В	8	7	8	8	8	9	48
П2 — А	9	8	9	9	9	10	54
П2 — В	9	8	9	9	9	10	54
П3 — А	9	9	9	9	9	10	55
П3 — В	9	8	9	9	9	10	54
П4 — А	9	9	9	9	9	10	55
П4 — В	9	8	9	9	9	10	54
П5 — А	9	9	9	9	9	10	55
П5 — В	9	8	9	9	9	10	54
П6 — А	9	9	9	9	9	10	55

Комментарий (Аннотация 1): Наиболее стабильные результаты дают промпты 4–5, так как принуждают модель явно учитывать область и терминологию. Промпт 1 иногда приводит к менее согласованным ключевым словам (например, anomaly search вместо anomaly detection).

6.2. Оценка переводов Аннотации 2

Вариант	Грамм.	Термин.	Естеств.	Стиль	Смысл	Идиомы	Итого/60
П1 — А	9	9	8	9	9	10	54
П1 — В	9	8	8	9	9	10	53
П2 — А	9	9	9	9	9	10	55
П2 — В	9	9	9	9	9	10	55
П3 — А	9	9	9	9	9	10	55
П3 — В	9	9	9	9	9	10	55
П4 — А	9	9	9	9	9	10	55
П4 — В	9	9	9	9	9	10	55
П5 — А	9	9	9	9	9	10	55
П5 — В	9	9	9	9	9	10	55
П6 — А	9	9	9	9	9	10	55

Комментарий (Аннотация 2): Различия между промптами 2–5 минимальны, так как исходный текст уже строго научный.

Существенный вклад дает уточнение терминов: customer churn, dummy variables, class imbalance, RandomOverSampler. Промпт 5 удобен тем, что заранее фиксирует неоднозначные слова («факторы» → predictors/features).

7. Выводы

7.1. Рекомендуемые варианты перевода

Аннотация 1: рекомендован вариант Промпт 5 — Модель А.

Аннотация 2: рекомендован вариант Промпт 5 — Модель А.

7.2. Цепочка промптов для выбранного варианта

Для Аннотации 1 (Промпт 5 — Модель А):

1) Этап 1 (анализ неоднозначных слов):

Внимательно проанализируй предоставленный русскоязычный текст и выдели отдельные слова, которые могут иметь несколько значений...

2) Этап 2 (уточненный перевод):

Используя таблицу, выполни перевод. Для следующих слов я выбираю:

- ["идентификации"] → detection
- ["отклонений"] → deviations
- ["потоках"] → streams
- ["средств"] → tools

Для Аннотации 2 (Промпт 5 — Модель А):

1) Этап 1 (анализ неоднозначных слов).

2) Этап 2 (уточненный перевод) с выбранными значениями:

- ["факторы"] → predictors
- ["фиктивные"] → dummy
- ["производительность"] → performance
- ["вклад"] → contribution

Дополнительные уточняющие промпты:

Дополнительные уточнения не использовались (0 уточнений), так как качество итоговых вариантов было удовлетворительным.

7.3. Обоснование выбора

Промпт 5 показал наилучшее качество и управляемость: этап анализа позволяет заранее зафиксировать перевод неоднозначных терминов и избежать смысловых сдвигов (например, «идентификация» как identification vs detection, «факторы» как factors vs predictors/features). Итоговый перевод сохраняет научный стиль, корректную терминологию, естественен для англоязычной научной аудитории и точно передает смысл исходного текста.