

Министерство образования Республики Беларусь

Учреждение образования  
БЕЛОРУССКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ИНФОРМАТИКИ И РАДИОЭЛЕКТРОНИКИ

УДК 004.421.4:004.67:005.5

**ПАСКРОБКА**  
Георгий Сергеевич

**СИСТЕМА ОБРАБОТКИ ОТЧЁТОВ IT-КОМПАНИИ НА ОСНОВЕ  
ТЕХНОЛОГИИ МАШИННОГО ОБУЧЕНИЯ**

**АВТОРЕФЕРАТ**  
диссертации на соискание степени  
магистра

по специальности 1-39 80 03 Электронные системы и технологии  
(профилизация «Компьютерные технологии проектирования  
электронных систем»)

Минск 2022

Работа выполнена на кафедре проектирования информационно-компьютерных систем учреждения образования «Белорусский государственный университет информатики и радиоэлектроники»

Научный руководитель: **РОЛИЧ Олег Чеславович**,  
кандидат технических наук, доцент кафедры проектирования информационно-компьютерных систем учреждения образования «Белорусский государственный университет информатики и радиоэлектроники»

Рецензент: **ПОЛОЗКОВ Юрий Владимирович**,  
кандидат технических наук, доцент, заведующий кафедрой «Программное обеспечение вычислительной техники и автоматизированных систем» учреждения образования «Белорусский национальный технический университет»

Защита диссертации состоится «26» апреля 2022 г. года в 14<sup>00</sup> часов на заседании Государственной экзаменационной комиссии по защите магистерских диссертаций в учреждении образования «Белорусский государственный университет информатики и радиоэлектроники» по адресу: 220013, Минск, ул. П.Бровки, 6, копр. 1, ауд. 415а, тел. 293-20-80, E-mail: kafpiks@bsuir.by

С диссертацией можно ознакомиться в библиотеке учреждения образования «Белорусский государственный университет информатики и радиоэлектроники».

## ВВЕДЕНИЕ

В настоящее время трудно представить рынок труда без сферы информационных технологий. Стремительно растущий потенциал *IT* обеспечивает столь же стремительное сокращение издержек в производственной сфере, способствует облегчению и улучшению уровня жизни, открывает все новые и новые возможности для людей.

Количество *IT*-компаний растет с каждым годом. На сегодняшний день в Беларуси насчитывается минимум 50 крупных *IT*-компаний, известных во всем мире, а также работающих на иностранный рынок.

В своей деятельности *IT*-компания должна не только ориентироваться на свое состояние в актуальный момент времени, но и уметь рационально оценить свое потенциальное состояние в будущем. Это необходимо ей не только для оценки и предотвращения потенциальных рисков в своей будущей деятельности, но и для достижения конкурентных преимуществ в соперничестве с другими компаниями.

Так как всякое управленческое решение по своей сути является прогнозным, то прогнозирование является одной из фундаментальных основ управленческой деятельности. Цель организационного прогнозирования – предвидение будущего состояния организации. Результатом этой деятельности является разработка комплекса прогнозов, которые представляют собой вероятностное суждение о будущем состоянии всей организации или ее частей.

В современном мире задача прогнозирования хорошо решается с помощью машинного обучения – науки об алгоритмах, которые сами настраиваются на данные. Многие крупные компании, занимающиеся продажей товаров и услуг, внедряют проекты машинного обучения в процессы обработки и анализа отчетов с целью построения прогнозов. Однако тема машинного обучения еще недостаточно изучена с точки зрения эффективности тех или иных алгоритмов относительно разных сфер бизнеса.

С целью сохранения конкурентоспособности на рынке информационных технологий и услуг актуальны задачи оперативной и интеллектуальной обработки отчетов *IT*-компаний для анализа бизнес-процессов, устранения их недостатков, поиска путей оптимизации и совершенствования работы, прогнозирования динамики деятельности. Задача эффективности применения различных моделей машинного обучения для оценки и прогнозирования основных показателей производительности *IT*-компаний малоизучена.

Все вышеизложенное определило направление диссертационной работы, в рамках которой рассматривается автоматизация обработки отчетов *IT*-компаний для прогнозирования динамики ее деятельности с использованием технологий машинного обучения.

## ОБЩАЯ ХАРАКТЕРИСТИКА РАБОТЫ

### **Актуальность темы исследования**

С целью сохранения конкурентоспособности на рынке информационных технологий и услуг актуальны задачи оперативной и интеллектуальной обработки отчетов ИТ-компаний для анализа бизнес-процессов, устранения их недостатков, поиска путей оптимизации и совершенствования работы, прогнозирования динамики деятельности.

В связи с вышесказанным, исследование методов и алгоритмов машинного обучения для обработки отчетов ИТ-компаний является актуальным.

### **Степень разработанности проблемы**

Исследование методов и алгоритмов машинного обучения для обработки отчетов ИТ-компаний выполнялось с использованием работ российских, белорусских и зарубежных ученых: Владимир Вьюгин, Георгий Кухарев, Педро Домингос, Хенрик Бринк, Джозеф Ричардс, Марк Феверолф, Себастьян Рашка и др.

Одним из недостатков исследований, представленных в современной технической литературе, является незначительное рассмотрение прикладных задач, направленных на практическую реализацию новых теоретических разработок в области эффективных моделей машинного обучения для прогнозирования на базе отчетности ИТ-компаний.

### **Цель и задачи исследования**

Целью диссертации является автоматизация обработки отчетов ИТ-компаний с прогнозированием динамики ее деятельности.

Поставленная цель работы определяет следующие основные задачи:

- 1 Классифицировать показатели деятельности и типы отчетов ИТ-компаний;
- 2 Проанализировать методы обработки отчетов и прогнозирования деятельности ИТ-компаний;
- 3 Разработать алгоритмы обработки отчетов и прогнозирования деятельности ИТ-компаний на основе технологии машинного обучения.
- 4 Проанализировать эффективность разработанных алгоритмов обработки отчетов и прогнозирования деятельности ИТ-компаний на основе технологий машинного обучения и сделать вывод об эффективности полученных моделей.

### **Область исследования**

Содержание диссертации соответствует образовательному стандарту высшего образования второй ступени (магистратуры) ОСВО 1-39 80 03-2019 специальности 1-39 80 03 Электронные системы и технологии (профилизация «Компьютерные технологии проектирования электронных систем»).

## **Теоретическая и методологическая основа исследования**

*Информационная база* исследования сформирована на основе литературы, открытой информации, технических нормативно-правовых актов, сведений из электронных ресурсов, а также материалов научных конференций и семинаров.

## **Научная новизна**

*Научная новизна* и значимость полученных результатов работы заключается в применении методов предсказания на базе технологии машинного обучения при прогнозировании деятельности ИТ-компаний.

*Теоретическая значимость* работы заключается в математической модели функционирования ИТ-компаний и способах оценки их деятельности по концентрируемым в отчетах компаний данным.

*Практическая значимость* диссертации состоит в модели анализа квартальных отчетов ИТ-компаний, осуществляющей прогнозирование потенциального состояния компании по основным показателям ее деятельности для достижения в будущем максимальных конкурентных преимуществ.

## **Основные положения, выносимые на защиту**

1 Модель обработки отчетов по критерию эффективности на базе матрицы путаницы, позволяющая повысить точность прогнозирования деятельности ИТ-компаний до 73,5 %.

2 Алгоритм прогнозирования деятельности ИТ-компаний методами линейного предсказания на основе машинного обучения с адаптивной импульсной характеристикой, динамически корректируемой матрицей путаницы.

3 Методика анализа алгоритмов и определения оптимального алгоритма машинного обучения в прогнозировании деятельности продуктовой ИТ-компаний по ее ежеквартальным отчетам, заключающаяся в оценке эффективности по основным показателям: отзыв, точность измерений,  $F$ -мера точности теста.

## **Публикации**

Изложенные в диссертации основные положения и выводы опубликованы в 7 печатных работах.

Общий объем публикаций по теме диссертации составляет 35 страниц.

## **Структура и объем работы**

Диссертация состоит из введения, общей характеристики работы, трех глав с краткими выводами по каждой главе, заключения, библиографического списка и приложений.

**В первой главе** проведен анализ моделей формирования и способов обработки отчетов ИТ-компаний. Выявлено, что для оценки эффективности ИТ-подразделений служат различные показатели, причем применение тех или иных показателей тесно связано с уровнем зрелости ИТ-подразделения. Основными отчетами о деятельности ИТ-службы являются: срез по текущему состоянию работ и выполнению задач ИТ-подразделения, отчет о выполнении Стратегических инициатив в области автоматизации, еженедельный отчет по всем направлениям деятельности ИТ-подразделения и ежеквартальный отчет службы. Набор данных отчетов является неотъемлемым инструментом в корректировке «курса» решаемых задач подразделением. Среди современных технологий, применяемых для анализа и обработки отчетов ИТ-компаний выделены *Data Mining* и *Knowledge Discovery in Databases (KDD)*.

**Во второй главе** описаны и проанализированы модели функционирования ИТ-компаний, а также методы модели обработки отчетов ИТ-компаний на базе технологий машинного обучения. Рассмотрены популярные алгоритмы машинного обучения, такие как линейная регрессия, логистическая регрессия, *LDA*, деревья принятия решений, наивный Байесовский классификатор, *KNN*, *LVQ*, *SVM*, случайный лес, *AdaBoost*. В качестве абстрактной ИТ-компания для внедрения технологии машинного обучения для анализа данных и построения прогнозов на базе отчетов в данной магистерской диссертации выбрана продуктовая ИТ-компания, занимающаяся проектированием комплексных клиент-серверных интернет-приложений. Компания имеет третий уровень зрелости по модели *CobIT* и ведет подробную отчетность о результатах своей деятельности.

**В третьей главе** разработан и проведен эксперимент по апробации десяти моделей машинного обучения для обработки ежеквартальных отчетов по 18 основным показателям деятельности продуктовой ИТ-компания. Для оценки эффективности моделей использована матрица путаницы. В ходе анализа матрицы путаницы для каждого из десяти алгоритмов, сделан вывод об эффективности полученных моделей.

**В приложении** представлены листинг кода, публикации автора, презентация, справка из системы «Антиплагиат».

Общий объем диссертационной работы составляет 141 страницы. Из них 77 страниц основного текста, 28 иллюстраций на 28 страницах, 3 таблицы на 3 страницы, библиографический список из 20 наименований на 2 страницах, список собственных публикаций соискателя из 8 наименований на 2 страницах, 4 приложения на 64 страницах.

## ОСНОВНОЕ СОДЕРЖАНИЕ РАБОТЫ

**Во введении** рассмотрено современное состояние проблемы оценки и прогнозирования основных показателей производительности ИТ-компания, а также дано обоснование актуальности темы.

**В общей характеристике работы** показана актуальность проводимых исследований, степень разработанности проблемы, сформулированы цель и задачи диссертации, обозначена область исследований, научная (теоретическая и практическая) значимость исследований, а также апробация работы.

**В первой главе** проведен анализ моделей формирования и способов обработки отчетов ИТ-компаний. Если руководство ИТ-компаний заинтересовано в таком мощном и эффективном союзнике как Информационные технологии для достижения требуемых результатов, поставленных задач перед бизнесом, то оно (руководство) обязано знать на каком уровне зрелости находится ИТ-подразделение и вовремя реагировать на выявленные отклонения. Целевой уровень зрелости ИТ процессов – «оптимизация», этап, когда создание новых бизнес-процессов, товаров и услуг, а также расширение границ бизнеса может и должно происходить с помощью информационных технологий.

Для оценки эффективности ИТ-подразделений служат различные показатели, причем применение тех или иных показателей тесно связано с уровнем зрелости ИТ-подразделения.

Основными отчетами о деятельности ИТ-службы являются:

- срез по текущему состоянию работ и выполнению задач ИТ-подразделения – отчет о плане/факте работ;

- отчет о выполнении стратегических инициатив в области автоматизации (может являться частью отчета о плане/факте);

- еженедельный отчет по всем направлениям деятельности ИТ-подразделения и ежеквартальный отчет службы.

Набор данных отчетов является неотъемлемым инструментом в корректировке «курса» решаемых задач подразделением.

Для автоматизации анализа отчетов используются две современные технологии: *Data Mining* и *Knowledge Discovery in Databases (KDD)*. *Data Mining* позволяет осуществлять оценивание, определение отклонений, прогнозирование (с помощью методов математической статистики, машинного обучения, нейронных сетей и др.).

**Во второй главе** описаны и проанализированы модели функционирования ИТ-компаний, а также методы обработки отчетов ИТ-компаний на базе технологий машинного обучения.

В зависимости от направления деятельности, выделяют следующие разновидности ИТ-компаний: продуктовая, *outsourcing*, *outstaffing*, консалтинговые, ИТ-отделы нетехнических компаний (например, банк).

Машинное обучение широко применяется для анализа данных и прогнозирования. Популярны алгоритмы машинного обучения: линейная регрессия, логистическая регрессия, *LDA*, деревья принятия решений, наивный Байесовский классификатор, *KNN*, *LVQ*, *SVM*, случайный лес, *AdaBoost*. Чтобы получить надежные результаты, разработчикам нужно отлично владеть математическим моделированием и статистическими принципами для подготовки надежного набора данных, что является важным условием успешного машинного обучения.

Успех проектов машинного обучения в промышленности определяется множеством факторов: слаженностью и опытом команды, пониманием данных и технологических процессов, оптимальным распределением ресурсов в ходе проекта и др. Четкое следование бизнес-цели проекта машинного обучения позволяет исследовать данные, строить гипотезы и проверять их, выбрав оптимальные прототипы моделей и спланировав поэтапное тестирование. Такой подход позволяет сэкономить ресурсы и на ранних этапах исключить проекты, не приносящие экономической выгоды.

В качестве абстрактной ИТ-компании для внедрения технологии машинного обучения для анализа данных и построения прогнозов на базе отчетов в данной магистерской диссертации выбрана продуктовая ИТ-компания, занимающаяся проектированием комплексных клиент-серверных интернет-приложений. Компания имеет третий уровень зрелости по модели *CobIT* и ведет подробную отчетность о результатах своей деятельности.

В третьей главе разработан и проведен эксперимент по апробации десяти моделей машинного обучения для обработки ежеквартальных отчетов по 18 основным показателям деятельности продуктовой ИТ-компании и проведен анализ полученных показателей.

Программная реализация алгоритмов машинного обучения выполнена на языке программирования *Python*. В качестве *GUI* для проекта использован *Jupyter Notebook*, широко используемый для машинного обучения и разработки приложений *data science*.

Алгоритм программы обработки отчетов ИТ-компании представлен на рисунке 1.

Шаг «Предварительная обработка данных» включает в себя следующие этапы:

- загрузка данных о цене, ежеквартальный отчет и сохранение как фрейма данных;
- очистка и обработка данных;
- объединение этих двух фреймов данных;
- расчет процентной разницы;
- экспорт данных в новый файл Excel.

Шаг «Создание метрик» – создание метрик в отношении процентного изменения цены. Есть 3 метрики:

- покупка (если самый высокий максимум и самый низкий минимум квартала увеличатся на 3% или более в следующем квартале);
- продажа (если самый низкий минимум и самый высокий максимум квартала снизятся на 3% или более в следующем квартале);
- удержание (если это не произойдет ни уменьшения, ни увеличения самого низкого минимума и самого высокого максимума в следующем квартале).





**Рисунок 1 – Алгоритм программы обработки отчетов IT-компании**

Для оценки эффективности моделей использовалась матрица путаницы. Матрица путаницы – это показатель успешности классификации, где классов два или более.

На рисунке 2 представлена матрица путаницы модели *AdaBoost* для прогнозирования на базе ежеквартальных отчетов продуктовой *IT*-компании.

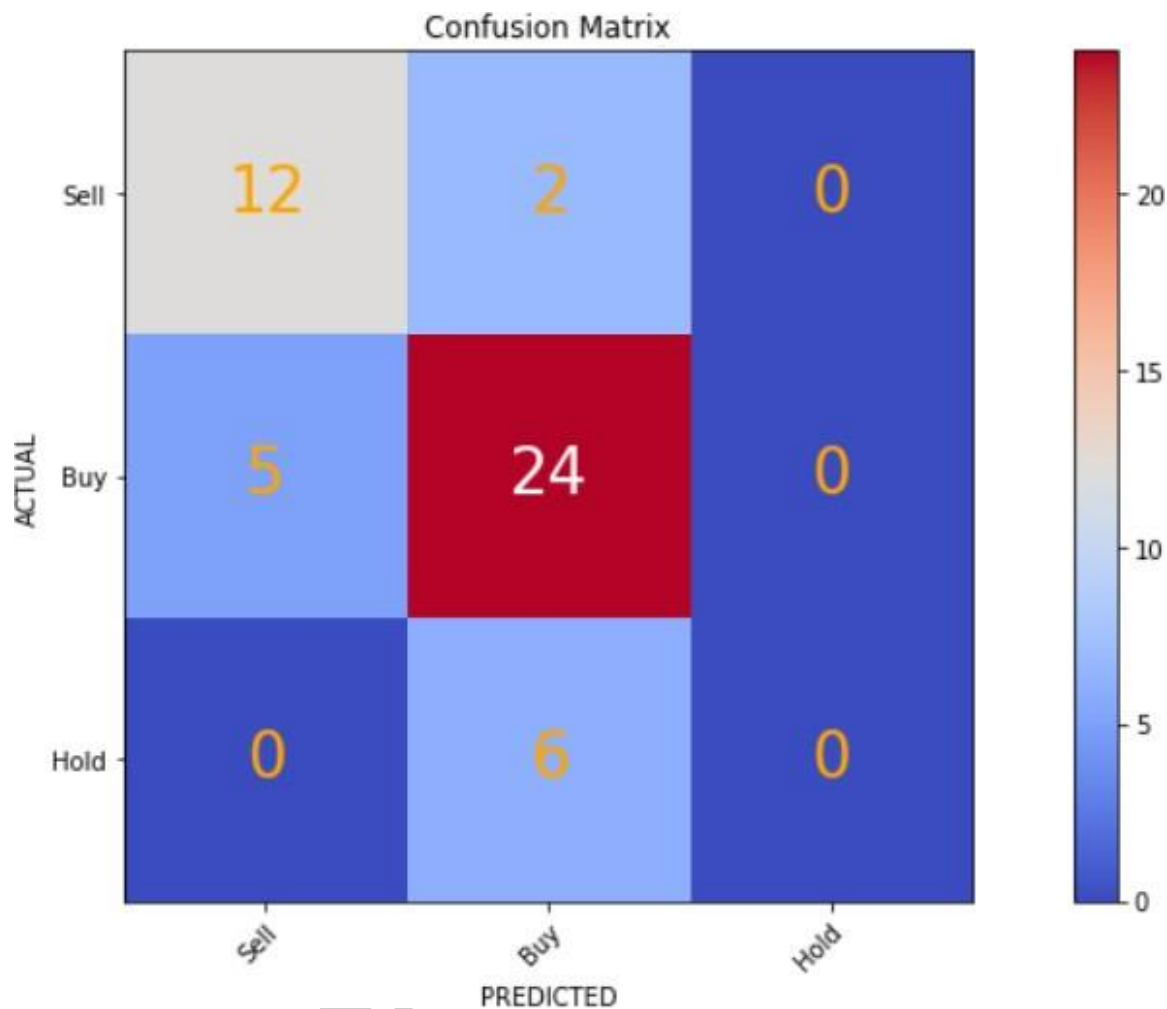


Рисунок 2 – Матрица путаницы модели *AdaBoost*

Вычисляем ключевые показатели эффективности модели *AdaBoost* для прогнозирования на базе ежеквартальных отчетов продуктовой *IT*-компании исходя из данной матрицы (формулы 1 – 7).

$$Accuracy = \frac{12 + 24 + 0}{12 + 2 + 0 + 5 + 24 + 0 + 0 + 6 + 0} = 73,5\% \quad (1)$$

$$Recall_{sell} = \frac{12}{12 + 2 + 0} = 85,7\% \quad (2)$$

$$Precision_{sell} = \frac{12}{12 + 5 + 0} = 70,6\% \quad (3)$$

$$F - measure_{sell} = \frac{2 * 0,857 * 0,735}{0,857 + 0,735} = 79,1\% \quad (4)$$

$$Recall_{buy} = \frac{24}{24 + 5 + 0} = 82,7\% \quad (5)$$

$$Precision_{buy} = \frac{24}{2 + 24 + 6} = 75\% \quad (6)$$

$$F - measure_{buy} = \frac{2 * 0,872 * 0,735}{0,872 + 0,735} = 79,8\% \quad (7)$$

Точность данной модели составляет 73,5%. *F*-мера метрик *sell* и *buy* составляет 79,1% и 79,8% соответственно. Данный алгоритм является наиболее эффективным для прогнозирования на базе ежеквартальных отчетов продуктовой *IT*-компания. Точность прогнозирования используя остальные девять моделей машинного обучения составляет менее 50% и, следовательно, не могут быть использованы для обработки отчетов *IT*-компания без совершенствования.

## ЗАКЛЮЧЕНИЕ

### Основные научные результаты диссертации

1 В ходе выполнения данного проекта изучены степени зрелости *IT*-компания по модели *CobIT*. Целевой уровень зрелости *IT*-процессов – «Оптимизация», этап, когда создание новых бизнес-процессов, товаров и услуг, а также расширение границ бизнеса может и должно происходить с помощью информационных технологий. Для оценки эффективности *IT*-подразделений служат различные показатели, причем применение тех или иных показателей тесно связано с уровнем зрелости *IT*-подразделения. Чем выше уровень зрелости компания, тем более эффективно ведется отчетность. Для автоматизации анализа отчетов используются две современные технологии: *Data Mining* и *Knowledge Discovery in Databases (KDD)*. *Data Mining* позволяет осуществлять оценивание, определение отклонений, прогнозирование [1, 4, 5, 6].

2 Машинное обучение широко применяется для анализа данных и прогнозирования. Популярные алгоритмы машинного обучения: линейная регрессия, логистическая регрессия, *LDA*, деревья принятия решений, наивный Байесовский классификатор, *KNN*, *LVQ*, *SVM*, случайный лес, *Ada-Boost* [1, 2, 7].

3 В качестве абстрактной ИТ-компании для внедрения технологии машинного обучения для анализа данных и построения прогнозов на базе отчетов в данной магистерской диссертации выбрана продуктовая ИТ-компания, занимающаяся проектированием комплексных клиент-серверных интернет-приложений. Компания имеет третий уровень зрелости по модели *CobIT* и ведет подробную отчетность о результатах своей деятельности [3].

4 В ходе выполнения данного проекта разработан и проведен эксперимент по апробации десяти моделей машинного обучения для обработки ежеквартальных отчетов по 18 основным показателям деятельности продуктовой ИТ-компании.

5 Для оценки эффективности моделей машинного обучения использовалась матрица путаницы. В ходе анализа матрицы путаницы для каждого из десяти алгоритмов, установлено, что лучше всего показал себя алгоритм *AdaBoost* (предсказания с точностью 73,5%). Точность прогнозирования используя остальные модели составляет менее 50% и, следовательно, не могут быть использованы для обработки отчетов ИТ-компании без соответствующей доработки.

#### **Рекомендации по практическому использованию результатов**

Полученные результаты могут быть использованы анализа и прогнозирования на базе отчетов любой продуктовой ИТ-компании.

#### **Возможности по дальнейшему развитию**

Результат данной работы представляет собой первую версию конечного продукта. В дальнейшем планируется внести следующие доработки:

1 Обучение моделей планируется проводить на большем объеме данных ИТ-компании, чтобы получить более точный анализ применимости соответствующих моделей машинного обучения.

2 Планируется проанализировать еженедельный отчет, а также отчеты по другим *KPI* компании. Планируется выполнить прогнозирование дополнительных зависимостей вместе с зависимостью изменения цены от ежеквартальных показателей.

3 Для оценки эффективности моделей машинного обучения планируется использовать также другие методы анализа (такие как кривая *ROC*, кривая *Precision-Recall* и др.) в совокупности с матрицей путаницы для получения более точных результатов.

По возможности, эксперимент планируется провести на реальных отчетах крупной ИТ-компании со штатом сотрудников не менее 1000 человек.

## СПИСОК ПУБЛИКАЦИЙ СОИСКАТЕЛЯ

### *Материалы конференций*

[1] Паскробка Г.С. ИСПОЛЬЗОВАНИЕ МАШИННОГО ОБУЧЕНИЯ ДЛЯ КЛАССИФИКАЦИИ МНОГОСТРАНИЧНЫХ ДОКУМЕНТОВ / Г.С. Паскробка // Молодой исследователь: вызовы и перспективы: сб. ст. по материалам ССXLVIII Международной научно-практической конференции «Молодой исследователь: вызовы и перспективы». – № 6(248). – М., Изд. «Интернаука», 2022.

[2] Паскробка Г.С. ТОП-5 АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОБРАБОТКИ ОТЧЕТОВ ИТ-КОМПАНИИ / Г.С. Паскробка // Молодой исследователь: вызовы и перспективы: сб. ст. по материалам ССLXII Международной научно-практической конференции «Молодой исследователь: вызовы и перспективы». – № 20(262). – М., Изд. «Интернаука», 2022.

[3] Паскробка Г.С. ОСОБЕННОСТИ ВНЕДРЕНИЯ ТЕХНОЛОГИИ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОБРАБОТКИ ОТЧЕТОВ ИТ-КОМПАНИИ / Г.С. Паскробка // Молодой исследователь: вызовы и перспективы: сб. ст. по материалам ССLXII Международной научно-практической конференции «Молодой исследователь: вызовы и перспективы». – № 20(262). – М., Изд. «Интернаука», 2022.

### *Статьи в электронных научных журналах*

[4] Паскробка Г.С. ПРИНЦИПЫ ОЦЕНКИ ЭФФЕКТИВНОСТИ ДЕЯТЕЛЬНОСТИ ИТ-КОМПАНИИ // Интернаука: электрон. научн. журн. 2022. № 20(243). URL: <https://internauka.org/journal/science/internauka/243>

[5] Паскробка Г.С. ВЛИЯНИЕ УРОВНЯ ЗРЕЛОСТИ ИТ-КОМПАНИИ НА ВЫБОР ИНСТРУМЕНТОВ ОТЧЕТНОСТИ // Интернаука: электрон. научн. журн. 2022. № 20(243). URL: <https://internauka.org/journal/science/internauka/243>

[6] Паскробка Г.С. МЕТОДЫ ОБРАБОТКИ И АВТОМАТИЗИРОВАННОГО АНАЛИЗА ОТЧЕТОВ ИТ-КОМПАНИЙ // Студенческий вестник: электрон. научн. журн. 2022. № 20(212). URL: <https://studvestnik.ru/journal/stud/herald/212>

[7] Паскробка Г.С. МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ОБРАБОТКИ ОТЧЕТОВ ИТ-КОМПАНИЙ // Студенческий вестник: электрон. научн. журн. 2022. № 20(212). URL: <https://studvestnik.ru/journal/stud/herald/212>

**РЭЗІЮМЭ**  
**Паскробка Георгій Сяргеевіч**  
**Сістэма апрацоўкі справаздач ІТ-кампаніі на аснове тэхналогіі**  
**машыннага навучання**

**Ключавыя словы:** аўтаматызацыя, машыннае навучанне, апрацоўка справаздач, прагназаванне, кіраванне ІТ-кампаніяй.

**Мэта працы:** аўтаматызацыя апрацоўкі справаздач ІТ-кампаніі з прагназаваннем дынамікі яе дзейнасці.

**Атрыманыя вынікі і іх навізна:** Выяўлена, што ўжыванне тых ці іншых паказчыкаў для ацэнкі эфектыўнасці ІТ-кампаніі цесна звязана з узроўнем яе сталасці (па мадэлі *CobIT*). Чым вышэй узровень сталасці кампаніі, тым больш эфектыўна вядзецца справаздачнасць. Выяўлены папулярныя алгарытмы машыннага навучання, якія выкарыстоўваюцца для аналізу справаздач: лінейная рэгрэсія, лагістычная рэгрэсія, *LDA*, дрэвы прыняцця рашэнняў, наіўны Баесаўскі класіфікатар, *KNN*, *LVQ*, *SVM*, выпадковы лес, *AdaBoost*. У якасці абстрактнай ІТ-кампаніі для ўкаранення тэхналогіі машыннага навучання для аналізу дадзеных і пабудовы прагнозаў на базе справаздач у дадзенай працы абрана прадуктовая ІТ-кампанія, якая займаецца праектаваннем комплексных кліент-серверных інтэрнэт-прыкладанняў. Кампанія мае трэці ўзровень сталасці па мадэлі *CobIT* і вядзе падрабязную справаздачнасць аб выніках сваёй дзейнасці. З улікам гэтага распрацаваны і праведзены эксперымент па апрабацыі дзесяці мадэляў машыннага навучання для апрацоўкі штоквартальных справаздач па 18 асноўных паказчыках дзейнасці прадуктовай ІТ-кампаніі. Для адзнакі эфектыўнасці мадэляў машыннага навучання выкарыстоўвалася матрыца блытаніны. У ходзе аналізу матрыцы блытаніны для кожнага з дзесяці алгарытмаў, устаноўлена, што лепш за ўсе паказаў сябе алгарытм *AdaBoost* (прадказанні з дакладнасцю 73,5%). Дакладнасць прагназавання выкарыстоўваючы астатнія мадэлі складае менш за 50% і, такім чынам, не могуць быць выкарыстаны для апрацоўкі справаздач ІТ-кампаніі без адпаведнай дапрацоўкі. Значнасць атрыманых вынікаў працы заключаецца ва ўжыванні метадаў лінейнага прадказання на базе тэхналогіі машыннага навучання пры прагназаванні дзейнасці ІТ-кампаніі.

**Вобласць ужывання:** сістэмы прагназавання на базе справаздач прадуктовай ІТ-кампаніі.

## РЕЗЮМЕ

Паскробка Георгий Сергеевич

### Система обработки отчетов ИТ-компаний на основе технологии машинного обучения

**Ключевые слова:** автоматизация, машинное обучение, обработка отчетов, прогнозирование, управление ИТ-компанией.

**Цель работы:** автоматизация обработки отчетов ИТ-компаний с прогнозированием динамики ее деятельности.

**Полученные результаты и их новизна:** Выявлено, что применение тех или иных показателей для оценки эффективности ИТ-компаний тесно связано с уровнем ее зрелости (по модели *CobIT*). Чем выше уровень зрелости компании, тем более эффективно ведется отчетность. Выявлены популярные алгоритмы машинного обучения, используемые для анализа отчетов: линейная регрессия, логистическая регрессия, *LDA*, деревья принятия решений, наивный Байесовский классификатор, *KNN*, *LVQ*, *SVM*, случайный лес, *AdaBoost*. В качестве абстрактной ИТ-компаний для внедрения технологии машинного обучения для анализа данных и построения прогнозов на базе отчетов в данной работе выбрана продуктовая ИТ-компания, занимающаяся проектированием комплексных клиент-серверных интернет-приложений. Компания имеет третий уровень зрелости по модели *CobIT* и ведет подробную отчетность о результатах своей деятельности. С учетом этого разработан и проведен эксперимент по апробации десяти моделей машинного обучения для обработки ежеквартальных отчетов по 18 основным показателям деятельности продуктовой ИТ-компаний. Для оценки эффективности моделей машинного обучения использовалась матрица путаницы. В ходе анализа матрицы путаницы для каждого из десяти алгоритмов, установлено, что лучше всего показал себя алгоритм *AdaBoost* (предсказания с точностью 73,5%). Точность прогнозирования используя остальные модели составляет менее 50% и, следовательно, не могут быть использованы для обработки отчетов ИТ-компаний без соответствующей доработки. Значимость полученных результатов работы заключается в применении методов линейного предсказания на базе технологии машинного обучения при прогнозировании деятельности ИТ-компаний.

**Область применения:** системы прогнозирования на базе отчетов продуктовой ИТ-компаний.

## SUMMARY

Georgiy Sergeevich Paskrobka

### IT company report processing system based on machine learning technology

**Keywords:** automation, machine learning, report processing, forecasting, IT company management.

**The object of study:** automation of the processing of reports of an IT company with forecasting the dynamics of its activities.

**The results and novelty:** It was revealed that the use of certain indicators to assess the effectiveness of an IT company is closely related to the level of its maturity (according to the CobIT model). The higher the level of maturity of the company, the more effectively reporting is carried out. Popular machine learning algorithms used for report analysis are identified: linear regression, logistic regression, LDA, decision trees, naive Bayes classifier, KNN, LVQ, SVM, random forest, AdaBoost. As an abstract IT company for the implementation of machine learning technology for data analysis and forecasting based on reports, in this work we have chosen an IT product company engaged in the design of complex client-server Internet applications. The company has the third level of maturity according to the CobIT model and maintains detailed reporting on the results of its activities. With this in mind, an experiment was developed and conducted to test ten machine learning models for processing quarterly reports on 18 key performance indicators of a product IT company. A confusion matrix was used to evaluate the performance of machine learning models. During the analysis of the confusion matrix for each of the ten algorithms, it was found that the AdaBoost algorithm performed best (predictions with an accuracy of 73.5%). The accuracy of forecasting using other models is less than 50% and, therefore, cannot be used to process reports of an IT company without appropriate refinement. The significance of the obtained results of the work lies in the application of linear prediction methods based on machine learning technology in predicting the activities of an IT company.

**Sphere of application:** forecasting systems based on reports of a product IT company.