



# АННОТАЦИЯ

Выпускная квалификационная работа на тему «Разработка системы классификации обращений абонентов в колл-центр ГК "Орион Телеком"» содержит 43 страницы текстового документа, 3 иллюстрации, 18 таблиц, 3 приложения, 30 использованных источников.

Цель данной работы — разработка системы классификации обращений абонентов в колл-центр группы компаний "Орион Телеком".

Разработка системы проводилась на языке программирования Python. Для создания классификатора использовался пакет инструментов машинного обучения scikit-learn. Для распознавания речи абонентов был применён сервис по распознаванию речи на русском языке Yandex SpeechKit.

Классификация обращений абонентов — актуальная проблема современного клиенто-ориентированного бизнеса, в частности телекоммуникационных компаний, в которых взаимодействие клиента и оператора осуществляется максимально оперативно посредством телефонного разговора. Для своевременного принятия управленческих решений, бизнесу требуется оперативная и точная статистика существующих проблем и поступающих обращений, чего человеческие ресурсы колл-центра самостоятельно предоставить не в состоянии. Применение автоматических систем распознавания речи и классификации текстов решает данную проблему и позволяет бизнесу упростить процесс сбора данных и сосредоточиться на анализе и принятии решений.

Результатом данной работы является набор приложений предназначенных для распознавания текста в аудиофайлах обращений абонентов, классификации данных обращений по классам и подклассам, а также записи результатов распознавания и классификации в корпоративную базу данных. Итоговая система была внедрена в внутреннюю инфраструктуру компании ООО «Орион Телеком».

Ключевые слова: КЛАССИФИКАЦИЯ ТЕКСТОВ, ОБРАБОТКА ЕСТЕСТВЕННОГО ЯЗЫКА, ТЕЛЕКОММУНИКАЦИОННЫЕ УСЛУГИ.

# ANNOTATION

The graduate qualification work on «The development of the classification system of subscribers' calls to the call center of "Orion Telecom"» contains 43 pages of a text document, 3 illustrations, 18 tables, 3 applications, 30 used sources.

The purpose of this work is the development of a system of classification of subscribers' calls to the call-center of the Orion Telecom group of companies.

The development of the system was done in the Python programming language. The machine learning package scikit-learn was used to create the classifier. The service of speech recognition in Russian Yandex SpeechKit was used to recognize the speech of the subscribers.

Subscribers' calls classification is an actual problem of modern client-oriented business, in particular of telecommunication companies, in which the interaction between a client and an operator is realized very quickly with a telephone conversation. For timely decision-making, business requires operational and accurate statistics on existing problems and incoming calls, which human resources call-center can not provide alone. The application of automatic speech recognition and text classification systems solves this problem and allows business to simplify the process of data collection and focus on analysis and decision making.

The result of this work is a set of applications designed to recognize a text in audio files of the subscribers' calls, to classify these calls into classes and subclasses, as well as to record the results of recognition and classification in the corporate database. The final system was implemented in the internal infrastructure of Orion Telecom Ltd.

Keywords: TEXT CLASSIFICATION, NATURAL LANGUAGE PROCESSING, TELECOMMUNICATION SERVICES.

# СОДЕРЖАНИЕ

Введение . . . . .	3
1 Постановка задач . . . . .	5
1.1 Постановка задачи распознавания аудио . . . . .	5
1.2 Постановка задачи классификации текстов . . . . .	5
1.3 Описание набора обучающей выборки . . . . .	6
1.4 Постановка цели работы . . . . .	7
2 Получение текстов . . . . .	8
2.1 Меры качества распознавания речи . . . . .	8
2.2 Системы распознавания речи . . . . .	10
2.3 Сравнение систем распознавания речи . . . . .	11
2.4 Архитектура приложения . . . . .	16
3 Классификация текстов . . . . .	17
3.1 Предобработка текстов . . . . .	17
3.1.1 Стемминг и лемматизация . . . . .	18
3.1.2 Распознавание именованных сущностей . . . . .	18
3.1.3 Токенизация текстов . . . . .	19
3.1.4 Удаление наиболее и наименее частотных токенов . . . . .	19
3.1.5 Векторизация . . . . .	19
3.1.6 Дедубликация . . . . .	20
3.2 Меры качества классификации . . . . .	21
3.3 Сравнение моделей классификации . . . . .	22
4 Результаты работы . . . . .	26
4.1 Описание архитектуры итоговой модели . . . . .	26
4.2 Результаты работы на тестовой выборке . . . . .	27
4.3 Описание микросервисной инфраструктуры . . . . .	29
Заключение . . . . .	31
Список использованных источников . . . . .	32
Приложения . . . . .	36
Приложение А: Типы и причины обращений . . . . .	36
Приложение Б: Результаты классификации по причинам обращений . . . . .	39
Приложение В: Акт о внедрении результатов магистерской диссертации . . . . .	43

## ВВЕДЕНИЕ

Классификация обращений абонентов — актуальная проблема современного клиенто-ориентированного бизнеса, в частности телекоммуникационных компаний, в которых взаимодействие клиента и оператора осуществляется максимально оперативно посредством телефонного разговора.

В связи с постоянным и интенсивным потоком обращений, вся деятельность колл-центров направлена исключительно на решение проблем абонентов, при этом операторы физически не успевают формировать отчёты о причинах обращения пользователей. Для бизнеса же крайне важна информация о том, какие причины чаще всего побуждают абонентов обратиться с жалобой — как для оперативного разрешения внезапно возникших проблем, так и для устранения долговременных неудобств, жалобы на которые поступают со стабильной периодичностью.

Помимо этого, классификация обращений абонентов требуется и для оценки качества работы сотрудников колл-центра — для каждого класса обращений у операторов есть заранее прописанный менеджментом «скрипт» — набор вопросов и фраз, которые обязательно нужно задать и произнести при разговоре с абонентом. И, поскольку для каждого класса обращений «скрипты» отличаются, для выбора верного «скрипта» и сопоставления с ним расшифровки фраз оператора, также требуется классификация поступающих обращений.

Наём дополнительных сотрудников, отвечающих за постоянное составление отчётов невыгоден в условиях минимизации затрат — помимо трат на заработные платы, таким сотрудникам также придётся опрашивать операторов на предмет наиболее популярных жалоб, что так же будет отнимать время операторов от непосредственной работы с клиентами.

В подобных условиях, системы автоматической классификации обращений пользователей, решают три основные проблемы: удовлетворяют потребность в постоянном и оперативном информировании руководства о причинах жалоб клиентов; снижают загруженность операторов колл-центра; и минимизируют расходы на внутренние процессы компании.

Кроме того, поскольку для проведения классификации обращения, ауди-

озапись разговора оператора с абонентом требуется транскрибировать (также программными средствами) полезным побочным результатом такой системы является полная расшифровка диалога с клиентом, которая в дальнейшем может применяться для решения совершенно иных управленческих проблем (например анализа токсичности речи абонента для выявления проблемных клиентов или сопоставление фраз оператора со скриптом).

**Целью** настоящей работы является разработка системы классификации обращений абонентов в колл-центр группы компаний "Орион Телеком".

Для достижения указанной цели были поставлены следующие **задачи**:

1. Анализ существующих систем распознавания речи и выбор наиболее оптимальной системы;
2. Разработка приложения, применяющего выбранную систему для распознавания записей звонков пользователей;
3. Разработка приложения для классификации распознанных текстов;
4. Проведение оценки качества классификации обращений на тестовой выборке и реальных данных.

Для разработки приложений использовались следующие программные **средства**:

1. Язык программирования Python версии 3.8
2. Пакет инструментов машинного обучения scikit-learn
3. СУБД ClickHouse
4. Брокер сообщений RabbitMQ
5. Контейнеризатор приложений Docker
6. Платформа для управления контейнеризованными сервисами Kubernetes

**Структура работы.** Выпускная квалификационная работа объёмом 43 страницы состоит из *Введения*, четырёх глав, *Заключения* и списка источников. Работа содержит 3 иллюстрации, 18 таблиц, 3 приложения, 30 использованных источников.

# 1 Постановка задач

В настоящей главе вводятся базовые определения и формальные постановки задач, описываются исходные данные, а также обосновывается постановка цели работы.

## 1.1 Постановка задачи распознавания аудио

Распознавание речи — процесс преобразования речевого сигнала в цифровую информацию (например, текстовые данные) [1]. В данной работе исходными данными являлись записи диалогов абонентов, обратившихся в колл-центр, и операторов, сотрудников колл-центра: одноканальные записи фраз оператора (снятые с микрофона оператора); одноканальные записи фраз абонента (снятые с микрофона телефона абонента); а также объединённые одноканальные записи всего диалога. Все аудиофайлы изначально записаны в формате WAV с битрейтом (скоростью потока) 128 кбит в секунду.

Для разработки собственной системы распознавания речи всегда требуется обучающая выборка транскрибированных аудиозаписей человеческой речи различных людей, например датасет Golos содержит 1 103 799 записей [2]. Создание подобного корпуса транскрибированных аудиозаписей сопряжено с невероятными материальными и временными затратами, а качественного корпуса транскрибированных записей звонков обнаружено не было, потому задача по распознаванию речи свелась к поиску подходящей существующей системы распознавания разговорной речи на русском языке. Подробнее об этом в *Главе 2*.

## 1.2 Постановка задачи классификации текстов

Имеется множество классов  $\mathcal{C} = \{c_1, \dots, c_{|\mathcal{C}|}\}$  и множество документов  $\mathcal{D} = \{d_1, \dots, d_{|\mathcal{D}|}\}$ . Неизвестна целевая функция-классификатор  $\Phi: \mathcal{C} \times \mathcal{D} \rightarrow \{0, 1\}$ . Необходимо построить классификатор  $\Phi'$ , максимально близкий к  $\Phi$ .

Имеется некоторая начальная коллекция размеченных документов  $\mathcal{X} \subset \mathcal{C} \times \mathcal{D}$ , для которых известны значения  $\Phi$ . Обычно её делят на «обучающую»

и «тестовую» части. Первая используется для обучения классификатора, вторая — для независимой проверки качества его работы.

Классификатор может выдавать точный ответ  $\Phi': \mathcal{C} \times \mathcal{D} \rightarrow \{0, 1\}$  или степень подобия  $\Phi': \mathcal{C} \times \mathcal{D} \rightarrow [0, 1]$  [3]. В текущей задаче требуется чёткий классификатор  $\Phi': \mathcal{C} \times \mathcal{D} \rightarrow \{0, 1\}$ . Поэтому можно переопределить искомую функцию как  $\Psi: \mathcal{D} \rightarrow \mathcal{C}$ .

Особенностью поставленной задачи является то, что классификация должна идти в два этапа: для целей менеджмента все обращения абонентов разделяются на восемь типов обращений (`request_type`), каждый из которых в свою очередь подразделяется на некоторое число причин обращений (`request_reason`). Все типы и причины обращений перечислены в *Приложении А*.

### 1.3 Описание набора обучающей выборки

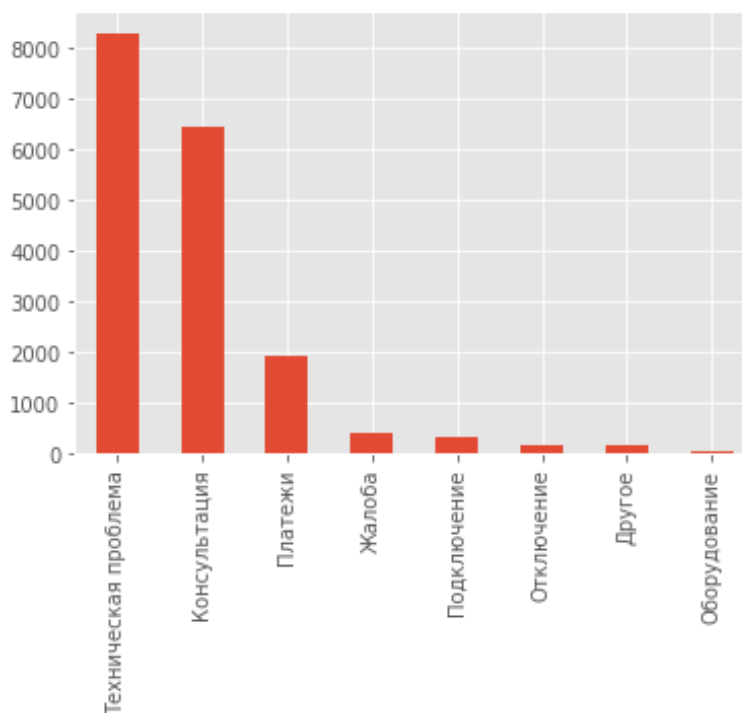


Рисунок 1 — Распределение документов по `request_type`

Для создания обучающей выборки, силами операторов колл-центра ГК «Орион Телеком» было размечено множество обращений пользователей — для каждого обращения операторы вручную проставляли тип и причину. В



результате разметки было получено 17 672 обращения, формирующих выборку, на базе которой и проводилась разработка и тестирование классификатора.

Очевидной проблемой данной выборки является явная несбалансированность классов. Из-за редкости некоторых классов, руководством компании было обозначено, что классификация по `request_type` «Другое» и «Оборудование» не обязана быть качественной.

#### 1.4 Постановка цели работы

Подводя итоги первой главы, можно перечислить требования к разрабатываемой системе:

1. Система должна использовать внешний сервис для транскрибирования аудиозаписи диалога абонента и оператора;
2. Система должна классифицировать полученные транскрипции диалогов по типам обращений, а затем по причинам обращений;
3. Система должна выводить данные в корпоративные базы данных для предоставления возможности менеджменту компании дальнейшей работы с ними;
4. Система должна быть эффективной настолько, насколько это возможно;

Таким образом, основной целью настоящей работы ставится разработка системы классификации обращений абонентов в колл-центр группы компаний «Орион Телеком», соответствующей вышеперечисленным требованиям.

Именно исходя из этих принципов в *Введении* был сформулирован список задач и поставлена изначальная цель.

## ЗАКЛЮЧЕНИЕ

Исходя из цели и задач, поставленных в данной работе, можно утверждать, что в ходе выполнения данной работы все поставленные задачи были выполнены, а цель достигнута:

1. Проведён анализ существующих систем распознавания речи и выбор наиболее оптимальной системы;
2. Разработано приложение, применяющее выбранную систему для распознавания записей звонков пользователей;
3. Разработано приложение для классификации распознанных текстов;
4. Проведена оценка качества классификации обращений на тестовой выборке и реальных данных.

В качестве итога работы разработано два микросервисных приложения, которые включены в программную инфраструктуру ГК «Орион Телеком» и в данный момент используются менеджментом (см. Приложение В).

Материалы данной диссертации были представлены на IV Всероссийской с международным участием научно-практической конференции студентов, аспирантов и работников образования и промышленности СУИТиММ 2022.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

- [1] Распознавание речи [Электронный ресурс] – Режим доступа: [https://ru.bmstu.wiki/Распознавание\\_речи](https://ru.bmstu.wiki/Распознавание_речи) (13.04.22).
- [2] *Karpov, N.* Golos: Russian Dataset for Speech Research [Электронный ресурс] - Режим доступа: <https://doi.org/10.48550/arXiv.2106.10161> (13.04.22) / N. Karpov, A. Denisenko, F. Minkin. — 2021.
- [3] Классификация текстов и анализ тональности [Электронный ресурс] – Режим доступа: [https://neerc.ifmo.ru/wiki/index.php?title=Классификация\\_текстов\\_и\\_анализ\\_тональности](https://neerc.ifmo.ru/wiki/index.php?title=Классификация_текстов_и_анализ_тональности) (13.04.22).
- [4] *Карпов, А. А.* Методология оценивания работы систем автоматического распознавания речи / А. А. Карпов, И. С. Кипяткова // *Информационные технологии автоматического анализа речи.* — 2012. — Vol. 55, no. 11. — Pp. 38–43.
- [5] *Levenshtein, V. I.* Binary codes capable of correcting deletions, insertions and reversals / V. I. Levenshtein // *Sov. Phys. Dokl.* — 1966. — Vol. 6. — Pp. 707–710.
- [6] *Khokhlov, Y.* Speech recognition performance evaluation for LVCSR system / Y. Khokhlov, N. Tomashenko // *Proc. of the 14th Intern. Conf. “Speech and Computer” SPECOM–2011, Kazan, Russia.* — 2011. — Pp. 129–135.
- [7] *Nghiên, Cuu Khoa Hoc.* Word Error Rate (WER) and Word Recognition Rate (WRR) with Python [Электронный ресурс] – Режим доступа: <https://holianh.github.io/portfolio/Cach-tinh-WER/>.
- [8] Unsupervised segmentation of words into morphemes — Morpho challenge 2005 Application to automatic speech recognition / M. Kurimo, M. Creutz, M. Varjokallio et al. // *Proc. Interspeech-2006, Pittsburgh, PA.* — 2006. — Pp. 1021–1024.
- [9] *Schlippe, T.* Grapheme-to-phoneme model generation for indo-european languages / T. Schlippe, S. Ochs, T. Schultz // *Proc. ICASSP-2012, Kyoto, Japan.* — 2012.

- [10] Accent modeling based on pronunciation dictionary adaptation for large vocabulary Mandarin speech recognition / C. Huang, E. Chang, J. Zhou, K. Lee // *Proc. Interspeech-2000, Beijing, China.* — 2000. — Pp. 818–821.
- [11] Uyghur morpheme-based language models and ASR / M. Ablimit, G. Neubig, M. Mimura S. et al. // *Proc. of the 10th IEEE Intern. Conf. on Signal Processing ICSP-2010, Beijing, China.* — 2010. — Pp. 581–584.
- [12] *Karpov, A.* Very large vocabulary ASR for spoken russian with syntactic and morphemic analysis / A. Karpov, I. Kipyatkova, A. Ronzhin // *Proc. Interspeech-2011, Florence, Italy.* — 2011. — Pp. 3161–3164.
- [13] Testing of the Speech Recognition Systems Using Russian Language Models / K. Aksyonov, D. Antipin, T. Afanaseva et al. // *Proceedings of the 5th International Young Scientists Conference on Information Technologies, Telecommunications and Control Systems 2018.* — 2018. — Vol. 2298.
- [14] Google Cloud Speech-to-Text [Электронный ресурс] – Режим доступа: <https://cloud.google.com/speech-to-text> (26.04.22).
- [15] Yandex.Speech Kit [Электронный ресурс] – Режим доступа: <https://cloud.yandex.ru/services/speechkit> (26.04.22).
- [16] *Zueva, N.* PocketSphinx , Audio speech recognition [Электронный ресурс] – Режим доступа: <https://drive.google.com/drive/folders/18tUAMхаQcm-M827Tvoo1uL0SKRFGxdQTP> (26.04.22).
- [17] Speech-to-Text (Russian) by SergeySHK [Электронный ресурс] – Режим доступа: <https://github.com/SergeyShk/Speech-to-Text-Russian> (26.04.22).
- [18] Обработка естественного языка [Электронный ресурс] – Режим доступа: [http://neerc.ifmo.ru/wiki/index.php?title=Обработка\\_естественного\\_языка](http://neerc.ifmo.ru/wiki/index.php?title=Обработка_естественного_языка) (09.05.22).
- [19] Словарь терминов, Стоп-слова [Электронный ресурс] – Режим доступа: [https://promopult.ru/library/Стоп\\_слова](https://promopult.ru/library/Стоп_слова) (11.05.2022).

- [20] Оценка качества в задачах классификации [Электронный ресурс] – Режим доступа: [http://neerc.ifmo.ru/wiki/index.php?title=Оценка\\_качества\\_в\\_задачах\\_классификации](http://neerc.ifmo.ru/wiki/index.php?title=Оценка_качества_в_задачах_классификации) (20.05.2022).
- [21] *Breiman, L.* Random Forests / L. Breiman // *Machine Learning*. — 2001. — Vol. 45. — Pp. 5–32.
- [22] *Breiman, L.* Bagging predictors / L. Breiman // *Machine Learning*. — 1996. — Vol. 24(2). — Pp. 123–140.
- [23] *Friedman, J.* Greedy Function Approximation: A Gradient Boosting Machine / J. Friedman // *The Annals of Statistics*. — 2001. — Vol. 29(5).
- [24] *Rabiner, Lawrence R.* A tutorial on hidden Markov models and selected applications in speech recognition / Lawrence R. Rabiner // *Proceedings of the IEEE*. — 1989. — Vol. 77(2). — Pp. 257–286.
- [25] *Mihaila, G.* GPT2 For Text Classification using Hugging Face Transformers [Электронный ресурс] – Режим доступа: <https://gmihaila.medium.com/fine-tune-transformers-in-pytorch-using-transformers-57b40450635> (23.05.2022). — 2020.
- [26] *Zadrozny, B.* Transforming classifier scores into multiclass probability estimates / B. Zadrozny, C. Elkan // *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*. — 2002. — Pp. 694–699.
- [27] Combining Text Classification and Hidden Markov Modeling Techniques for Structuring Randomized Clinical Trial Abstracts / Rong Xu, Kaustubh Supekar, Yang Huang et al. // *AMIA Annu Symposium Proceedings*. — 2006. — Pp. 824–828.
- [28] *Рысьмятова, А.* Классификация текстов [Электронный ресурс] – Режим доступа: [http://www.machinelearning.ru/wiki/images/a/a9/Rysmyatova\\_report.pdf](http://www.machinelearning.ru/wiki/images/a/a9/Rysmyatova_report.pdf) (23.05.2022). — 2015.

- [29] Stochastic Gradient Descent [Электронный ресурс] – Режим доступа: <https://scikit-learn.org/stable/modules/sgd.html> (24.05.2022).
- [30] *Яшневский, Н.* Автоматизация обработки клиентских обращений [Электронный ресурс] – Режим доступа: <https://habr.com/ru/company/X5Group/blog/598919/> (24.05.2022). — 2022.

## ПРИЛОЖЕНИЯ

### Приложение А: Типы и причины обращений

Тип обращения	Причина обращения
Подключение	Новое подключение / заключение договора
Подключение	Подключение доп. услуги
Подключение	Подключение доп. адреса/точки
Подключение	Перезаключение договора
Подключение	Смена тарифного плана
Отключение	Приостановка обслуживания
Отключение	Отключение доп. услуги
Отключение	Отключение доп. адреса/точки
Отключение	Расторжение договора
Отключение	Повышение стоимости АП
Консультация	Смена тарифного плана
Консультация	Выдача логина и пароля
Консультация	По сайту / личному кабинету
Консультация	По оплате
Консультация	Уточнение по адресу подключения
Консультация	Уточнение по срокам подключения
Консультация	Уточнение по тарифному плану
Консультация	Уточнение по заявке
Консультация	По оборудованию
Консультация	Прочее
Платежи	Текущие платежи
Платежи	Авансовый платёж
Платежи	Перерасчёт
Платежи	Ошибочный платёж
Платежи	Запрос детализации
Платежи	Вопрос по балансу
Платежи	Непонятное списание со счёта

Платежи	Подключение кредита
Платежи	Новогодний бонус
Техническая проблема	Вызов инженера
Техническая проблема	Заявка в службу 911
Техническая проблема	Регистрация нового оборудования (1 линия)
Техническая проблема	Решение проблемы (2 линия)
Оборудование	Выдача
Оборудование	Продажа
Оборудование	Возврат (14 дней)
Оборудование	Возврат гарантийного оборудования
Оборудование	Возврат по аренде
Оборудование	Замена (14 дней)
Оборудование	Замена гарантийного оборудования
Оборудование	Замена по аренде
Жалоба	Не устраивает качество выполнения заявки
Жалоба	Не устраивает качество услуг
Жалоба	Перебои соединения/обрывы связи
Жалоба	Жалобы на скорость интернета
Жалоба	Неудобный сайт / личный кабинет
Жалоба	Невыгодный тарифы
Жалоба	Некомпетентность сотрудников
Жалоба	Долгое решение проблемы
Жалоба	Долгий дозвон в call-центр
Жалоба	Долгое устранение неполадок
Жалоба	Принудительный перевод на тариф
Жалоба	Низкое качество оборудования
Жалоба	Недружелюбное отношение сотрудников
Жалоба	Очереди в центрах обслуживания клиентов
Жалоба	Нет оповещений о сбоях / тех. работах
Жалоба	Проблемы с файлообменником
Жалоба	Изменение сетки вещания каналов



Жалоба	Размещение рекламной продукции
Другое	Незавершённый звонок
Другое	Благодарность абонента
Другое	Ошибочное обращение
Другое	Не указано

Таблица 10 – Типы обращений и соответствующие причины обращений

## Приложение Б: Результаты классификации по причинам обращений

Причина обращения	Precision	Recall	$F_1$ -мера	Количество
Вызов инженера	0.91	0.88	0.89	125
Заявка в службу 911	0.63	0.71	0.67	17
Регистрация нового оборудования (1 линия)	0.90	0.92	0.91	108
Решение проблемы (2 линия)	0.97	0.97	0.97	578
Итого	0.85	0.87	0.86	828

Таблица 11 — Результаты классификации по request\_reason «Техническая проблема»  
(Accuracy = 0.94)

Причина обращения	Precision	Recall	$F_1$ -мера	Количество
Смена тарифного плана	0.14	1.00	0.25	1
Выдача логина и пароля	0.09	0.20	0.13	5
По оборудованию	0.70	0.59	0.64	64
По оплате	0.87	0.84	0.86	268
По сайту / личному кабинету	0.43	0.39	0.41	38
Прочее	0.67	0.71	0.69	79
Уточнение по адресу подключения	0.45	0.62	0.53	8
Уточнение по заявке	0.85	0.86	0.86	136
Уточнение по срокам подключения	0.00	0.00	0.00	1
Уточнение по тарифному плану	0.60	0.59	0.60	44
Итого	0.48	0.58	0.50	644

Таблица 12 — Результаты классификации по request\_reason «Консультация» (Accuracy = 0.75)

Причина обращения	Precision	Recall	$F_1$ -мера	Количество
Авансовый платёж	0.00	0.00	0.00	0
Вопрос по балансу	0.92	0.81	0.87	150
Запрос детализации	0.00	0.00	0.00	0
Непонятное списание со счёта	0.00	0.00	0.00	0
Новогодний бонус	0.95	0.95	0.95	19
Ошибочный платёж	0.00	0.00	0.00	0
Подключение кредита	0.61	0.61	0.61	18
Текущие платежи	0.00	0.00	0.00	4
Итого	0.31	0.30	0.30	191

Таблица 13 — Результаты классификации по request\_reason «Платежи» (Accuracy = 0.79)

Причина обращения	Precision	Recall	$F_1$ -мера	Количество
Долгое решение проблемы	0.00	0.00	0.00	5
Долгое устранение неполадок	0.38	0.43	0.40	7
Жалобы на скорость интернета	0.60	0.75	0.67	4
Не устраивает качество выполнения заявки	0.00	0.00	0.00	1
Не устраивает качество услуг	0.00	0.00	0.00	0
Невыгодный тарифы	0.00	0.00	0.00	0
Некомпетентность сотрудников	0.00	0.00	0.00	0
Низкое качество оборудования	0.00	0.00	0.00	3

Перебои соединения/обрывы связи	0.86	0.63	0.73	19
Итого	0.20	0.20	0.20	39

Таблица 14 — Результаты классификации по request\_reason «Жалоба» (Accuracy = 0.46)

Причина обращения	Precision	Recall	$F_1$ -мера	Количество
Новое подключение / заключение договора	1.00	0.78	0.88	23
Перезаключение договора	0.33	1.00	0.50	1
Подключение доп. адреса/точки	0.60	1.00	0.75	3
Подключение доп. услуги	0.50	0.67	0.57	3
Итого	0.61	0.86	0.67	30

Таблица 15 — Результаты классификации по request\_reason «Подключение» (Accuracy = 0.80)

Причина обращения	Precision	Recall	$F_1$ -мера	Количество
Отключение доп. адреса/точки	0.00	0.00	0.00	1
Отключение доп. услуги	1.00	0.50	0.67	2
Приостановка обслуживания	1.00	0.80	0.89	5
Расторжение договора	0.82	1.00	0.90	9
Итого	0.70	0.57	0.61	17

Таблица 16 — Результаты классификации по request\_reason «Отключение» (Accuracy = 0.82)

Причина обращения	Precision	Recall	$F_1$ -мера	Количество
-------------------	-----------	--------	-------------	------------

Благодарность абонента	1.00	1.00	1.00	1
Незавершённый звонок	1.00	0.75	0.86	8
Ошибочное обращение	0.75	1.00	0.86	6
Итого	0.92	0.92	0.90	15

Таблица 17 — Результаты классификации по request\_reason «Другое» (Accuracy = 0.87)

Причина обращения	Precision	Recall	$F_1$ - мера	Коли- чество
Возврат по аренде	0.00	0.00	0.00	0
Замена гарантийного оборудо- вания	0.00	0.00	0.00	1
Продажа	1.00	1.00	1.00	1
Итого	0.33	0.33	0.33	2

Таблица 18 — Результаты классификации по request\_reason «Оборудование» (Accuracy = 0.50)

# Приложение В: Акт о внедрении результатов магистерской диссертации

ООО «Орион телеком»  
Юридический адрес: 660017, г. Красноярск, ул. Ленина 113, оф. 100,  
тел./факс: (391) 2-000-000  
ИНН 2466220319  
КПП 246601001  
ОГРН 1092468017729  
Расчетный счёт № 40702810804000022096  
СИБИРСКИЙ Ф-Л ПАО "ПРОМСВЯЗЬБАНК", Г. НОВОСИБИРСК  
к/с 30101810500000000816  
БИК 045004816  
Генеральный директор Немировский Максим Станиславович,  
действует на основании Устава

УТВЕРЖДАЮ

Косоганов Дмитрий Андреевич  
(руководитель, директор)

ООО «Орион телеком»  
(наименование предприятия)

Косоганов Д.А.  
(подпись) (Ф.И.О.)

« 17 » июля 20 22 г.



АКТ

## о внедрении результатов магистерской диссертации

на тему Разработка системы классификации обращений абонентов в колл-центр  
ГК "Орион телеком"  
(наименование выполненной диссертации)

по направлению 01.04.02 Прикладная математика и информатика  
(код и наименование)

по образовательной программе 01.04.02.06 Прикладная математика в  
гуманитарных и социально-экономических науках  
(код и наименование)

выполненную Жарковым Михаилом Сергеевичем  
(Ф.И. О. магистранта)

Материалы магистерской диссертации Жаркова М. С. внедрены в деятельность ООО «Орион телеком».

В диссертации описана архитектура сервисов Speech Recognizer и Call Classifier, применяемых в компании для создания расшифровок диалогов абонентов и операторов, и классификации полученных расшифровок.

Результаты классификации используются при анализе поступающих обращений для оперативного принятия управленческих решений.

Министерство науки и высшего образования РФ  
Федеральное государственное автономное  
образовательное учреждение высшего образования  
«СИБИРСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Институт математики и фундаментальной информатики  
Кафедра высшей и прикладной математики

**УТВЕРЖДАЮ**

Заведующий кафедрой

 Мысливец С. Г.

«17» июня 2022 г.

**МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ**  
**РАЗРАБОТКА СИСТЕМЫ КЛАССИФИКАЦИИ ОБРАЩЕНИЙ**  
**АБОНЕНТОВ В КОЛЛ-ЦЕНТР ГК "ОРИОН ТЕЛЕКОМ"**

Направление 01.04.02 Прикладная математика и информатика


Магистерская программа 01.04.02.01 Математическое моделирование

Руководитель

  
16.06.22


доцент, кандидат физико-математических наук В.Р. Куликов

Выпускник

  
16.06.22

М.С. Жарков

Нормоконтролер

  
23.06.22.

Т.Н. Шипина

Красноярск 2022