МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего образования
«ТЮМЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

ИНСТИТУТ МАТЕМАТИКИ И КОМПЬЮТЕРНЫХ НАУК
Кафедра программного обеспечения

РЕКОМЕНДОВАНО К ЗАЩИТЕ В ГЭК
И ПРОВЕРЕНО НА ОБЪЕМ
ЗАИМСТВОВАНИЯ
Заведующий кафедрой
д.п.н., профессор
И.Г. Захарова
29 июня 2018 г.

МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ

РАЗРАБОТКА СЕРВИСА ДЛЯ АНАЛИЗА РИСКОВ ИТ-ПРОЕКТОВ

02.04.03. Математическое обеспечение и администрирование информационных систем
Магистерская программа «Высокопроизводительные вычислительные системы»

Выполнила работу
Студентка 2 курса
очной формы обучения
Матвеева
Оксана
Сергеевна

Научный руководитель
к.т.н., доцент
Воробьева
Марина
Сергеевна

Консультант
Старший преподаватель
Воробьев
Артем
Максимович

Рецензент
к.п.н., доцент
Плотоненко
Юрий
Анатольевич

Тюмень-2018

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ

ИТ-проекты являются наиболее сложными и дорогостоящими при автоматизации деятельности предприятия и сопряжены с различными рисками.

Риск – вероятное событие, которое влияет на ход реализации проекта и может привести к изменению стоимости, сроков или качества продукта проекта.

Одним из важных рисков для ИТ-проектов является несвоевременное оказания ИТ-услуг. Оказание данных услуг регулируется соглашением об уровне сервиса (Service Level Agreement).

Service Level Agreement (соглашение об уровне сервиса) – формальное соглашение между Заказчиком услуги и Исполнителем, содержащее описание услуги, права и обязанности сторон и согласованный уровень качества предоставления данной услуги.

Когда пользователь сталкивается с какой-либо проблемой (неисправностью, сбоем, просто неумением), то рассчитывает получить квалифицированную помощь в работе с приобретенной им услугой или продуктом. При этом его интересует максимально быстрое разрешение проблемы.

Бизнес-процесс обработки и регистрации обращения пользователя осуществляется следующим образом:

Пользователь звонит или отправляет письмо по электронной почте диспетчеру с описанием проблемы. Специалист диспетчерской службы анализирует обращение и назначает сектор, ответственный за решение данной проблемы. После чего регистрирует обращение на сайте «Журнал заявок».

Из-за большого потока заявок или некорректного описания проблемы в обращении не всегда удается верно определить сектор, ответственный за решение данной проблемы. Вследствие этого решить обращение в установленные метрики времени не удается.

Целью данной работы является автоматизация процесса обработки и распределения обращений пользователей корпоративных систем.

Разрабатываемая система позволит не только проанализировать обращения, но и определить сектор для назначения данного обращения.

Актуальность разработки системы заключается в том, что система позволит ускорить время реагирования на запросы пользователей, повысит эффективность и оперативность решения проблемы, увеличит количество обрабатываемых обращений. Сократив время на реагирование и обработку обращений, можно избежать риска возникновения финансовых штрафов, которые возникают при несоблюдении условий соглашения.

ГЛАВА 1. АНАЛИЗ ПРЕДМЕТНОЙ ОБЛАСТИ

1.1. КОНТРОЛЬ SLA

В качестве решения одного из рисков был рассмотрен процесс предоставления ИТ-услуг и контроль над качеством их исполнения.

Данный процесс регулируется соглашением об уровне сервиса.

Service Level Agreement (соглашение об уровне сервиса) – формальное соглашение между Заказчиком услуги и Исполнителем, содержащее описание услуги, права и обязанности сторон и согласованный уровень качества предоставления данной услуги.

В данном документе прописаны все сроки выполнения и обработки заявок, поступающих от пользователей. Для регистрации и мониторинга обращений пользователей используется информационная система «Журнал заявок».

ИС «Журнал заявок» - приложение, предоставляющие набор инструментов для управления услугами поддержки ИТ, предоставления отчётности и улучшения всех процессов, связанных с обслуживанием инфраструктуры ИТ.

Жизненный цикл заявки представлен на Рисунке 1.1. и включает несколько этапов.

Рисунок 1.1. Жизненный цикл заявки

1. Регистрация – это время формирования заявки по обращению Пользователя и направление по электронной почте уведомления о факте регистрации с указанием следующей информации:
   - номер заявки;

- дата и время регистрации заявки;
- тип обращения;
- категория изменения;
- приоритет заявки;
- описание обращения пользователя;
- срок выполнения заявки.

2. Реакция – это время, в течение которого Пользователь получает уведомление о приеме обращения в работу, и в течение которого Исполнитель связывается с Пользователем для уточнения деталей обращения.
3. Решение – это время, в течение которого обращение должно быть решено.

Количество нарушений сроков исполнения запросов не должно превышать порогового значения в 3% от всех поступивших запросов за отчетный месяц. Несоблюдение данных метрик времени может привести к финансовым потерям.

Матрица определения зон ответственности и времени по этапам обработки обращения (для сервисной линии DVLP) представлена в Таблице 1.1.

Таблица 1.1. Матрица зон ответственности и времени по этапам обработки обращения

Уровень обслуживания	*Регистрация ИТ-обращения, час.*	*Принятие ИТ-обращения в работу, час.*	*Диагностика ИТ-обращения, час.*	*Разрешение ИТ-обращения, час.*
Профессионал	Телефон – по звонку; Эл. почта – 0,25 часа	1	до 4 часов	
Профессионал	Телефон – по звонку;	1	до 4 часов	

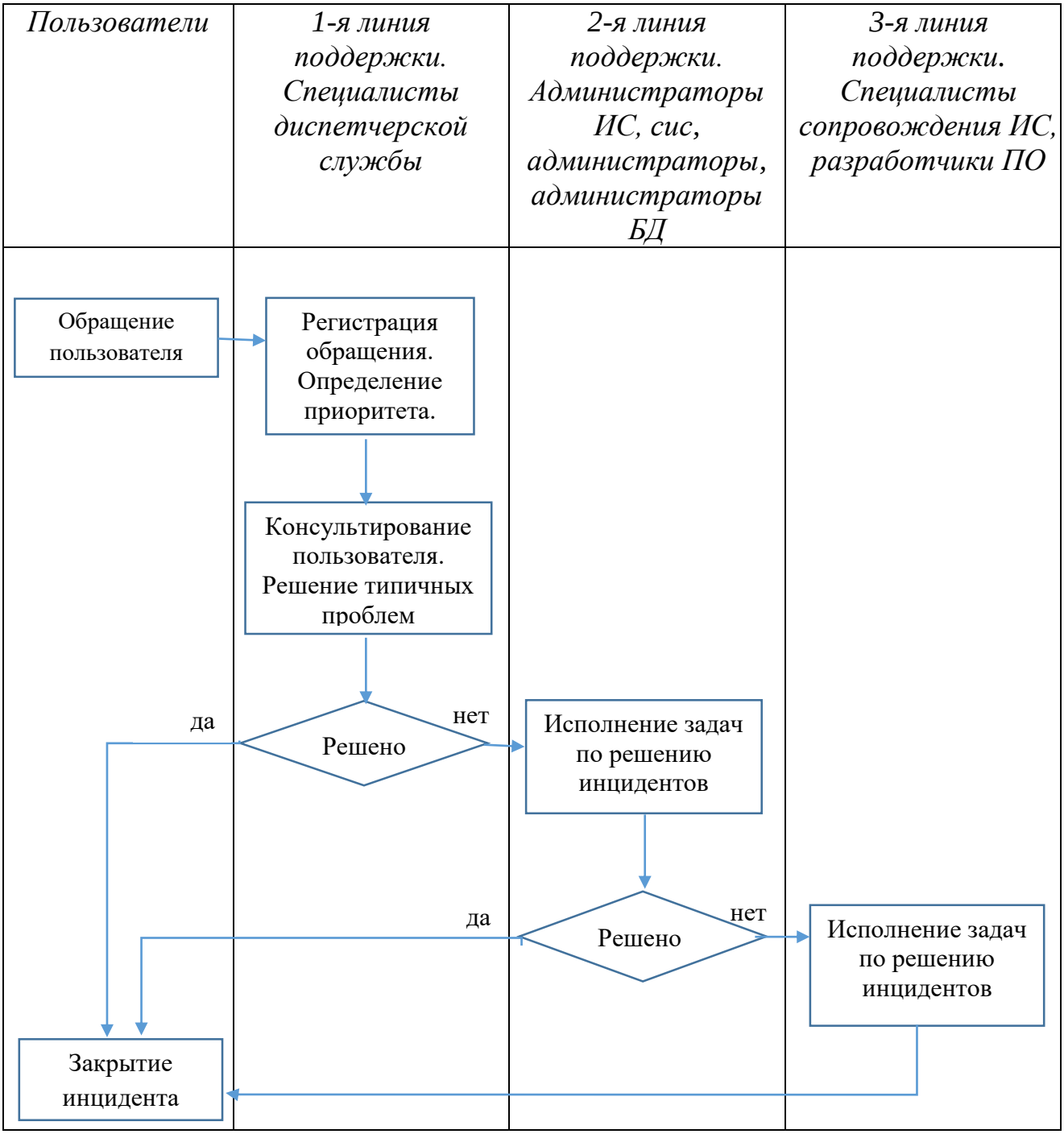
Уровень обслуживания	*Регистрация ИТ-обращения, час.*	*Принятие ИТ-обращения в работу, час.*	*Диагностика ИТ-обращения, час.*	*Разрешение ИТ-обращения, час.*
	Эл. почта – 0,25 часа			
Стандарт	Телефон – по звонку; Эл. почта - 0,25 часа	2	до 12 часов	
Эконом	Телефон – по звонку; Эл. почта - 0,25 часа	3	до 24 часов	

1.2. АЛГОРИТМ РЕШЕНИЯ ИНЦИДЕНТОВ

Для того чтобы правильно классифицировать обращение на линии поддержки, специалисты диспетчерской службы опираются на матрицу эскалаций.

Алгоритм эскалации обращения представлен в таблице 1.2.

Таблица 1.2. Схема эскалации обращения



Определив линию поддержки, необходимо определить сектор, ответственный за решение проблемы.

Информация по назначению заявок содержится в базе данных системы «ЦДС - Журнал заявок». На основе этих данных можно верно классифицировать обращения по подразделениям.

Для того чтобы обращение можно было зарегистрировать на сайте журнала заявок необходимо определить услугу, которую оказывает подразделение по данному обращению.

Для распределения обращения по подразделениям, назначения услуги и категории необходимо использовать методы классификации текста.

1.3. ОБЗОР МЕТОДОВ КЛАССИФИКАЦИИ ТЕКСТА

Формально постановку задачи классификации текста можно описать следующим образом.

Имеется множество документов D = {d1, d2...d3} и множество классов C = {c1, c2…c3}.

Неизвестная целевая функция Ф: D*C -> {0,1} задается формулой (1):

$$\Phi(d_j|ci) = \begin{cases} 0, \text{если } d_j \notin ci, \\ 1, \text{если } d_j \in ci \end{cases} \quad (1)$$

Необходимо найти функцию Ф’, приближенной к Ф.[2,8]

В данной работе рассмотрены методы классификации текса: метод Байеса, метод опорных векторов и метод k-ближайших соседей.

1. Наивный Байесовский классификатор

Метод Байеса относится к вероятностным методам классификации. Данный алгоритм предполагает, что наличие какого-либо признака не связано с наличием какого-либо другого признака. [2,4]

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)} \quad (2)$$

где P(c|x) – апостериорная вероятность класса при определенном значении признака X.

P(c) – априорная вероятность класса

P(x|c) – вероятность значения признака в классе

P(x) – априорная вероятность значения признака.

Байесовский классификатор использует оценку апостериорного максимума, для определения наиболее вероятного класса.

$$c\ map = argmax\ \frac{P(x|c)P(c)}{P(x)} \qquad (3)$$

В данном методе документ рассматривается как набор слов, вероятности которых условно не зависят друг от друга. Следовательно, условная вероятность документа:

$$P(x|c) = P(w1|c) * P(w2|c) * \dots P(wn|c) = \prod_{i=1}^{n} P(wi|c) \qquad (4)$$

$$c = argmax\ P(c) \prod_{i=1}^{n} P(wi|c) \qquad (5)$$

2. Метод опорных векторов

Данный метод применяется для задач классификации и регрессионного анализа [2,4]

Идея метода заключается в том, чтобы перевести вектора в пространство с более высокой размерностью и найти такую разделяющую гиперплоскость, которая будет иметь максимальный зазор в этом пространстве.

Оптимальная разделяющая гиперплоскость для метода опорных векторов строится на точках из двух классов. Ближайшие к гиперплоскостям точки называются опорными векторами.

Для задач многоклассовой классификации применяется подход One-vs-all.

3. Метод k-ближайших соседей

Метод K-ближайших соседей – один из методов решения задачи классификации.

Предполагается, что имеется какое-то количество объектов с точной классификацией. Нужно определить правило, позволяющее отнести новый объект к одному из возможных классов. Классы заранее известны.

В основе k-NN лежит следующее правило: объект считается принадлежащим тому классу, к которому относится большинство его

ближайших соседей. Под «соседями» понимаются объекты, близкие к исследуемому в том или ином смысле.

Данный метод не требует обучения. Документ, для которого надо определить класс, сравнивается со всеми документами из обучающей выборки и находится расстояние – косинус угла между документами.

$$p(d, d\text{в}) = \cos(d, d\text{в}) \quad (6)$$

После чего выбираются документы, ближайшие к d и для каждого класса вычисляется релевантность по формуле:

$$CSV(d, ci) = \sum \cos(d, d\text{в}) \quad (7)$$

Класс с наибольшей релевантностью относится к данному документу [2,4].

ГЛАВА 2. КЛАССИФИКАЦИЯ ДОКУМЕНТОВ

2.1. ОПИСАНИЕ МОДЕЛИ КЛАССИФИКАЦИИ

Классификацию обращений пользователей можно разделить на 3 этапа [8]:

1. Индексация обращения – построение числовой модели текста, в виде вектора слов и их веса в документе.
2. Построение и обучение классификатора – применение методов классификации текста.
3. Оценка качества классификации – оценка классификатора по таким критериям, как полнота, точность и численная оценка качества.

Описание последовательности построения классификатора представлено на рисунке 2.1.1.

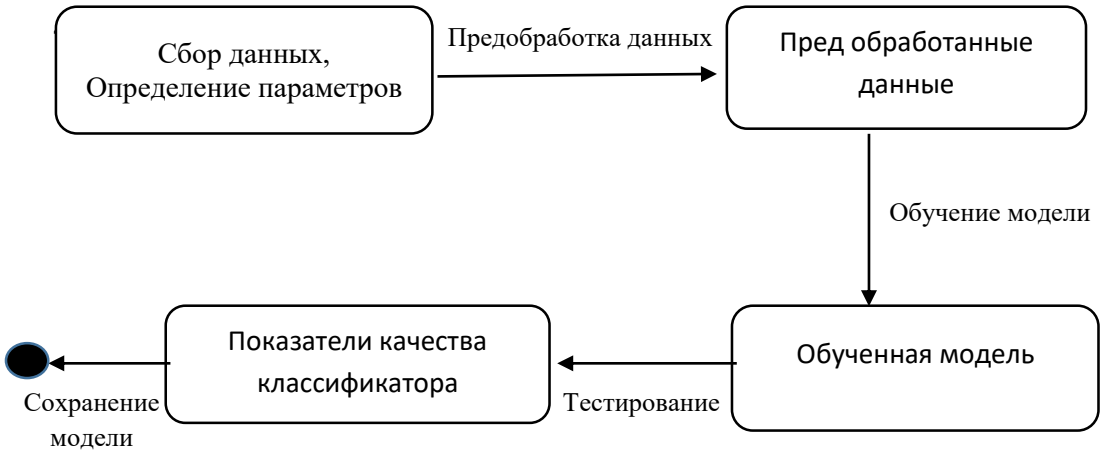


Рисунок 2.1.1. Последовательность создания классификатора

2.2. ОПИСАНИЕ ПАРАМЕТРОВ

Для того, чтобы верно классифицировать обращения пользователей по группам, надо определить параметры, которые влияют на выбор подразделения. Данные параметры представлены в таблице 2.2.1.

Таблица 2.2.1. Параметры для классификации обращения

Номер	*Параметр*
1	Организация
2	Пользователь
3	Описание
4	Группа исполнителей

Параметр «Группа исполнителей» - классы, на которые следует распределить обращения.

Параметр «Описание» - описание заявки.

Параметр «Пользователь» - пользователь системы. В зависимости от приоритета пользователя или его должности, обращения выполняют определенные подразделения.

Параметр «Организация» - название организации, в которой работает пользователь. В зависимости от организации обращения выполняют определенные подразделения.

Для регистрации заявки на сайте журнала заявок необходимо указать категорию обращения и наименование вида услуги, которая оказывается пользователю по его обращению. Параметры необходимые для классификации вида услуг представлены в таблице 2.2.2. Параметры необходимые для классификации категории обращения представлены в таблице 2.2.3.

Таблица 2.2.2. Параметры для классификации вида услуг

Номер	*Параметр*
1	Описание
2	Группа исполнителей
3	Услуга

Параметр «Услуга» - наименование услуги, которая будет оказана пользователю. Для каждого подразделения есть несколько видов услуг, по определенным типам обращений.

Таблица 2.2.3. Параметры для классификации категории

Номер	*Параметр*
1	Описание
2	Категория

Параметр «Категория» - тип обращения. Содержит 3 типа: инцидент, запрос на обслуживание, запрос на изменение.

2.3. ПРЕДОБРАБОТКА ДАННЫХ

После определения параметров классификации, необходимо выполнить предобработку текста. Для обработки текста на русском языке были использованы библиотеки PyMorphy2 и NLTK [10].

Предобработка текста включает:

- токенизацию
- удаление стоп-слов
- нормализацию слов

После предобработки текста, все документы необходимо представить в виде вектора R^n, где n - размерность вектора, соответствует количеству уникальных слов в корпусе.

Для преобразования коллекции документов в матрицу подсчета терминов была использована библиотека [10,14]:

from sklearn.feature_extraction.text import CountVectorizer.

После построения вектора слов, необходимо определить вес слова в документе.

2.4. ОПРЕДЕЛЕНИЕ ВЕСА ТЕРМИНОВ

Для определения веса терминов была использована библиотека [10,14]:

from sklearn.feature_extraction.text import TfidfVectorizer.

С помощью данной библиотеки коллекция документов преобразовывается в матрицу функций TF-IDF.

TF (term frequency — частота терма) — отношение числа вхождения некоторого терма к общему количеству термов документа. Таким образом, оценивается важность терма ti в пределах отдельного документа dj.

Пусть ft_{ij} — число вхождений терма t_i в документ d_j. Тогда частота терма определяется как:

$$\mathrm{TF}(\mathrm{ti}, \mathrm{dj}) = \frac{\mathrm{ftij}}{\sum_i \mathrm{ftij}} \qquad (8)$$

где $0 \le i \le |T|$, $0 \le j \le |D|$

IDF (inverse document frequency — обратная частота документа) — инверсия частоты, с которой некоторое слово встречается в документах коллекции. Учёт IDF уменьшает вес широкоупотребительных слов. Для каждого уникального слова в пределах конкретной коллекции документов существует только одно значение IDF.

$$\mathrm{TF}(\mathrm{ti}, \mathrm{dj}) = \log \frac{|\mathrm{D}|}{|(\mathrm{di} \supset \mathrm{ti})|} \qquad (9)$$

где |D| — количество документов в коллекции,

$|(di \supset ti)|$ — количество документов, в которых встречается ti (когда $ft_i \neq 0$) , $0 \le i \le |T|$.

TF-IDF — статистическая мера, используемая для оценки важности слова в контексте документа, являющегося частью коллекции документов или корпуса. Вес некоторого слова пропорционален количеству употребления этого слова в документе и обратно пропорционален частоте употребления слова в других документах коллекции [8].

$$wij = TF(\mathrm{t}i, dj) * IDF(\mathrm{t}i, D) \qquad (10)$$

2.5. ИСПОЛЬЗУЕМЫЕ МОДЕЛИ ОБУЧЕНИЯ

Для реализации алгоритмов классификации была использована библиотека языка Python - SciKit-Learn [14].

SciKit-Learn — пакет, предоставляющий эффективные версии множества распространенных алгоритмов [5].

Использование API для классификации текста с помощью библиотеки Scikit-Learn включает следующие шаги:

1. Выбор класса модели с помощью импорта соответствующего класса из библиотеки Scikit-Learn.

2. Компоновка данных в матрицу признаков

3. Обучение модели на своих данных посредством вызова метода fit() экземпляра модели.

4. Применение модели к новым данным:

* в случае машинного обучения с учителем метки для неизвестных данных обычно предсказывают с помощью метода predict();

* в случае машинного обучения без учителя выполняется преобразование свойств данных или вывод их значений посредством методов transform() или predict().

Используя данную библиотеку, был применен метод классификации текста: наивный Байесовский классификатор [16]:

Для реализации классификатора необходимо подключить библиотеку:

from sklearn.naive_bayes import MultinomialNB.

Класс MultinominalNB – используется для многоклассовой классификации.

Обучение модели классификатора и получения прогноза:

clf=MultinomialNB().fit(X_train, target)

predicted = clf.predict(X_train)

где X_train- коллекция документов из выборки

Target – классы, на которые следует распределить данные.

2.6. ОЦЕНКА КАЧЕСТВА КЛАССИФИКАТОРА

Для оценки работы классификатора используется несколько метрик: численная оценка алгоритма, точность и полнота [2,6,7].

Численная оценка качества (Accuracy) – доля, верно распределенных документов по классам.

$$\text{Accuracy} = \frac{P}{N} \quad (11)$$

где P- количество верно распределенных документов

N- Размер обучающей выборки.

Точность (precision) – доля документов, принадлежащих данному классу относительно всех документов в этом классе.

Полнота (recall) – доля документов, принадлежащих классу относительно всех документов этого класса в выборке.

Описание расчета для оценки качества представлено в таблице 2.3.1.

Таблица 2.3.1. Описание расчета для оценки качества классификатора

Категория		*Экспертная оценка*	
		Положительная	*Отрицательная*
Оценка системы	Положительная	TP	FP
	Отрицательная	FN	TN

TP – истинно-положительное решение

TN – истинно-отрицательное решение

FP – ложно-положительное решение

FN – ложно-отрицательное решение

$$\text{precision} = \frac{TP}{TP + FP} \quad (12)$$

$$\text{recall} = \frac{TP}{TP + FN} \quad (13)$$

F- мера - объединяет оценки о точности и полноте классификатора.

$$\text{F} = 2 \times \frac{precision \times recall}{precision + recall} \quad (14)$$

Пример оценки качества классификатора

Выборка: 100 обращений.

Класс А– 30 сообщений
Класс Б- 70 сообщений

Обработав сообщения, классификатор для класса А определил 20 сообщений, 16 из которых относятся к классу А.

Следовательно, TP=16, FN=14 и FP =4. Тогда, для класса А точность классификатора составляет 4/5 (80% положительных решений правильные).

Полнота классификатора равна 16/30 (классификатор нашел 53% документов класса А)

ГЛАВА 3. ОПИСАНИЕ РАЗРАБАТЫВАЕМОЙ СИСТЕМЫ

3.1. ТРЕБОВАНИЯ К СИСТЕМЕ

Целью работы является автоматизация процесса обработки заявок пользователей, с целью предотвращения наступления риска несвоевременного оказания услуг.

В связи с этим была поставлена задача - реализовать систему для интеллектуального распознавания и распределения обращений пользователей корпоративных систем, с помощью которой будет производиться анализ обращения и выдаваться рекомендация ответственного подразделения.

Разрабатываемая система должна удовлетворять следующим требованиям:

- Система должна быть реализована с применением web-технологий для доступа с любого устройства.
- Система должна использовать методы классификации текста для анализа обращений.
- Система должна обрабатывать текст, введенный пользователем и в короткий срок выдавать результат, в виде рекомендации ответственного подразделения.
- В случае, когда невозможно определить подразделение, система должна выдавать сообщение с просьбой уточнения проблемы или перевода на специалиста диспетчерской службы.
- Необходимо реализовать страницу с формой для ввода обращения пользователей и вывода результата.
- Система должна иметь доступ к базе данных Service Desk.
- Необходимо реализовать функционал, с помощью которого информация будет сохранена в базу данных журнала заявок, после чего заявка сформируется на портале ServiceDesk.
- Система должна быть протестирована на реальных данных.
- Результат тестирования системы должен составлять не менее 80% верно распределенных обращений.

– Система должна иметь интуитивно понятный, удобный и простой интерфейс.

Разрабатываемая система должна предоставлять следующий функционал:

1. Обработка обращения

После описания проблемы пользователем, должна выполняться обработка и анализ полученных данных, с помощью методов классификации текста, и отнесение обращения к какому-либо классу в соответствии с вероятностями слов для данного обращения.

2. Выдача рекомендаций

После того как текст будет отнесен к классу с большей вероятностью, должна выдаваться рекомендация, которая включает название подразделения, к которому относится решение данной проблемы. Если вероятность меньше 0,95, то пользователю выдается сообщение с просьбой уточнения проблемы или перевода на специалиста диспетчерской службы.

3. Регистрация обращения пользователя на сайте журнала заявок

После обработки обращения и выдачи рекомендации по обращению, пользователь может самостоятельно зарегистрировать заявку на сайте, нажав на кнопку. После чего, полученные данные будут сохранены в базе данных, и заявка отобразится на сайте.

3.2. ИСПОЛЬЗУМЫЕ ПО И ТЕХНОЛОГИИ

Для разработки системы был использован язык программирования Python. Среда разработки Sublime Text 3.

В качестве веб-сервера был использован микрофреймворк Flask.

Интерфейс системы написан на html с подключением скриптов Bootstrap для разметки страницы и оформления веб-компонент.

3.3. АРХИТЕКТУРА СИСТЕМЫ

Архитектуру системы можно разделить на 3 уровня: бизнес-логика, уровень приложения, уровень развертывания.

На уровне бизнес логики показаны возможные действия пользователей системы по ролям. К действиям пользователя относится введение обращения и регистрация заявки.

На уровне приложения показана программная реализация. В данном модуле происходит сбор и обработка данных, построение классификатора.

На уровне развертывания показаны технологии, которые будут использоваться. В данном модуле веб-сервис обрабатывает запрос пользователя и выдает результат в виде рекомендации по назначению подразделения, на основе обученного классификатора.

Архитектура системы представлена на рисунке 3.3.1.

3.4. ОПИСАНИЕ ДАННЫХ

В процессе обучения были разработаны 2 скрипта на языке Python и шаблон веб-страницы Index.http. Описание скриптов приведено в таблице 3.4.1.

Таблица 3.4.1. Описание скриптов, разрабатываемой системы

Название	*Описание*
Classification. py	Содержит методы построения и обучения классификатора, а также оценки качества построенной модели.
Main_app	Инициализация сайта

Код класса Main_app представлен в Приложении 3.

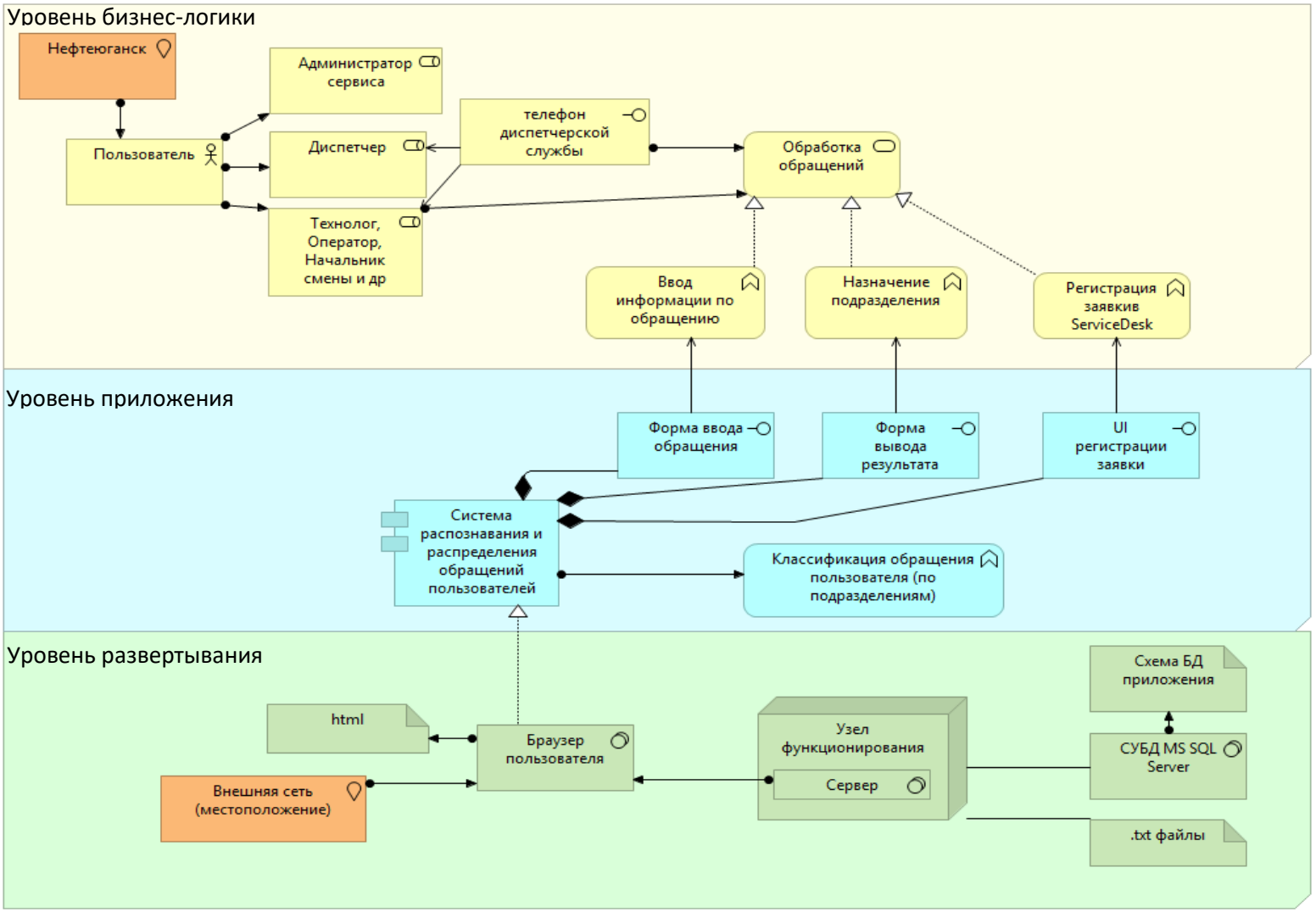


Рисунок 3.3.1. Архитектура системы

В скрипте Classification.py выполняется загрузка файла, в котором содержится информация по зарегистрированным обращениям пользователей. После извлекаются параметры для классификации, выполняется предобработка некоторых параметров, создается модель классификатора и выполняется оценка качества классификации. На основе обученного классификатора определяется группа исполнителей, услуга и категория для нового обращения.

Описание методов в скрипте Classification.py приведено в таблице 3.4.2. Код методов приведен в приложении 2.

Таблица 3.4.2. Описание методов, разрабатываемой системы

Название	*Описание*
Class_NB()	Извлечение информации по обращениям с excel-файла. Выбор параметров для создания модели: группа исполнителей, описание, пользователь, организация, категория, услуга. Реализация метода наивного Байеса для классификации текста. Оценка качества классификации.
Analyzer_text()	Анализ и предварительная обработка текста, введенного пользователем
Get_class ()	Определение группы исполнителей для обращения, введенного пользователем.

ГЛАВА 4. РЕЗУЛЬТАТ РАБОТЫ СИСТЕМЫ

4.1. ВХОДНЫЕ ДАННЫЕ

Входные данные представляют собой файл формата .xls. Файл размером 12Мб, содержит 12 414 строк и 25 столбцов. Строки — это количество обращений, которые были зарегистрированы с начала текущего месяца. Столбцы — это параметры, которые описывают обращение.

На рисунке 4.1.1. представлен excel-файл, содержащий информацию по обращениям, зарегистрированных на сайте «ЦДС-Журнал заявок».

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Номер	Статус	Категория	Статус согласования	Степень влияния	Конфигурационный элемент	Марка	Инв.номер	Услуга	Приоритет	Пользователь	Организация	Адрес	Телефон	Создана	Группа исполнителей	Исполнитель	Срок	Выполнена	Описание
2	1175489	Закрыта	Запрос на обслуживание	Согласовано	Низкое (на 1 персону)	NC-8181	Lenovo ThinkCentre M91p [4524PB3]	14254927	Комплексное обслуживание стационарного компьютера	Низкий-ИТ	Шарифуллин Рустам Шамилевич	ЮНГ- Управление повышения производит. резервуаров			05.05.2017 14:31	Системные администраторы здания П26		06.03.2018 17:02	05.03.2018 10:41	Предоставить 3 кабеля для подключения ПК к ИБП 40057853
3	1175506	Закрыта	Запрос на обслуживание	Согласовано	Низкое (на 1 персону)	NC-9154	HP Pro 3400 MT [LH201EA]	14263538	Комплексное обслуживание стационарного компьютера	Низкий-ИТ	Саидов Омар Алиевич	ЮНГ- Управление повышения производит. резервуаров			05.05.2017 15:07	Системные администраторы здания П26		11.04.2018 10:59	05.03.2018 10:46	Необходимы 2 шнура для подключения мониторов к ИБП инв. 40055362
4	1294641	Закрыта	Модернизация оборудования	Отказано	Низкое (на 1 персону)	Нет единого сервисного номера				Низкий-ИТ	Гуртов Алексей Борисович	ЮНГ- Управление скважинных технологий и супервайз			16.01.2018 14:33	Системные администраторы здания НКН		07.03.2018 12:18	05.03.2018 10:33	Выдать оборудование - Мобильный ПК (ноутбук)
5	1296496	Закрыта	Запрос на обслуживание	Согласовано	Низкое (на 1 персону)	NP-3715	HP Color LaserJet 5550	14212515	СЦ Обеспечение РМ и ЗИП Оргтехника	Средний	Хисматуллин Артем Маратович	ЮНГ- Управление ГРР,ресурсной базы и лицензирован			19.01.2018 10:50	Сервисный центр		07.03.2018 14:30	05.03.2018 14:26	Заменить картридж черный
6	1305771	Закрыта	Инцидент	Согласовано	Низкое (на 1 персону)	NC-7210	HP Pro 3400 MT [LH127EA]	14268712	Услуги сервисного центра	Низкий-ИТ	Цивковский Игорь Николаевич	ЮНГ- Управление информационных технологий			05.02.2018 10:45	Системные администраторы УДНГ ЮР, Юг. регион		05.03.2018 15:29	05.03.2018 15:07	Диагностика ПК
7	1306157	Закрыта	Модернизация оборудования	Согласовано	Низкое (на 1 персону)	Нет единого сервисного номера			Системное обслуживание стационарного компьютера	Низкий	Зарипов Александр Сергеевич	ООО Красноярск-Строй Инжиниринг			05.02.2018 17:01	Системные администраторы здания 7-49		05.03.2018 16:41	05.03.2018 15:43	Выдать оборудование - монитор
8	1306160	Закрыта	Модернизация оборудования	Согласовано	Низкое (на 1 персону)	Нет единого сервисного номера			Системное обслуживание стационарного компьютера	Низкий	Зарипов Александр Сергеевич	ООО Красноярск-Строй Инжиниринг			05.02.2018 17:05	Системные администраторы здания 7-49		05.03.2018 16:40	05.03.2018 15:42	Выдать оборудование - Стационарный ПК
9	1306172	Закрыта	Модернизация оборудования	Согласовано	Низкое (на 1 персону)	Нет единого сервисного номера			Системное обслуживание стационарного компьютера	Низкий	Писарева Олеся Александровна	ООО Красноярск-Строй Инжиниринг			05.02.2018 17:14	Системные администраторы здания 7-49		05.03.2018 16:26	05.03.2018 15:41	Выдать оборудование - Стационарный ПК
10	1306177	Закрыта	Модернизация оборудования	Согласовано	Низкое (на 1 персону)	Нет единого сервисного номера			Системное обслуживание стационарного компьютера	Низкий	Наумов Станислав Владимирович	ООО Красноярск-Строй Инжиниринг			05.02.2018 17:19	Системные администраторы здания 7-49		05.03.2018 16:35	05.03.2018 15:42	Выдать оборудование - Стационарный ПК
11	1306184	Закрыта	Модернизация оборудования	Согласовано	Низкое (на 1 персону)	Нет единого сервисного номера			Системное обслуживание стационарного компьютера	Низкий	Рогозинский Максим Юрьевич	ООО Красноярск-Строй Инжиниринг			05.02.2018 17:33	Системные администраторы здания 7-49		05.03.2018 16:21	05.03.2018 15:40	Выдать оборудование - Стационарный ПК
12	1306363	Закрыта	Модернизация оборудования	Согласовано	Низкое (на 1 персону)	Нет единого сервисного номера			Системное обслуживание стационарного компьютера	Низкий-ИТ	Досчанова Эльвира Радиковна	ЮНГ- Управление по договорной работе			06.02.2018 0:51	Системные администраторы здания 7-49		05.03.2018 16:23	05.03.2018 15:40	Выдать оборудование Стационарный ПК
13	1309315	Закрыта	Постановка на обслуживание	Согласовано	Низкое (на 1 персону)	Нет единого сервисного номера			Сопровождение СЭД БОСС-Референт	Низкий	Асмендиярова Рима Рифовна	ООО РН-Учет ТУЦ в г. Уфа			09.02.2018 21:46	ОИС нормативно-справочной информации		16.03.2018 10:44	05.03.2018 10:36	1607-13-2174

Рисунок 4.1.1. Выгрузка информации с сайта ЦДС-Журнал заявок

4.2. ЗАГРУЗКА ДАННЫХ

Для загрузки и обработки данных использовались средства библиотеки pandas. Для работы с файлами формата .xls была добавлена библиотека import xlrd, xlwt.

Данные были разделены на обучающую и тестовую выборку в процентном отношении 60/40.

Необходимые параметры для построения и обучения классификатора — это «Категория», «Услуга», «Пользователь», «Организация», «Группа исполнителей», «Описание».

```
rb = xlrd.open_workbook('Servicecall_new.xls', formatting_info = True)
sheet = rb.sheet_by_index(0)
val = sheet.row_values(0)[0]
print('Выборка состоит из: ')

n=round(sheet.nrows*0.6)

category = [sheet.row_values(rownum)[3] for rownum in range(n)]
service = [sheet.row_values(rownum)[9] for rownum in range(n)]
user = [sheet.row_values(rownum)[12] for rownum in range(n)]
company = [sheet.row_values(rownum)[13] for rownum in range(n)]
address = [sheet.row_values(rownum)[14] for rownum in range(n)]
valsgroup = [sheet.row_values(rownum)[17] for rownum in range(n)]
valsdescrp = [sheet.row_values(rownum)[21] for rownum in range(n)]

request_df = pd.DataFrame({'company': company,'users': user,'description': valsdescrp,'name': valsgroup})
service_df = pd.DataFrame({'service': service,'description': valsdescrp,'name': valsgroup,'newname':''})
category_df = pd.DataFrame({'description': valsdescrp,'category': category})
```

Рисунок 4.2.1. Чтение файла и извлечение параметров

Полученная обучающая выборка состоит из 7448 обращений. Количество подразделений в данной выборке - 78. В таблице 4.2.1. представлено распределение количества обращений по группам.

Таблица 4.2.1. Распределение количества обращений по группам исполнителей

Группа исполнителей	*Количество обращений*
Группа первой линии поддержки пользователей HP	223

Группа исполнителей	*Количество обращений*
Группа сопровождения ВКС	26
Группа сопровождения ИС Инфраструктура	66
Отдел информационной безопасности	116
Отдел информационных систем ФХД	127
Отдел подготовки данных	89
Отдел системной инженерии	169
Отдел телекоммуникационных радиотехнологий	54
Поддержка ИС DIRECTUM	317
Проектная группа сопровождения ЭПОС	292
Сектор связи ИТ	3
Сектор сопровожд. систем нефтедобычи и разработки	161
Сектор сопровождения оперативных систем (ССОС)	137
Сектор сопровождения систем технол. мониторинга	102
Сервисный центр	374
ЦАП-1 (Юганский регион)	225
ЦАП-2 (Приобский регион)	43
ЦАП-3 (Майский регион)	348
ЦАП-4 (Мамонтово)	326
ЦАП-5	202
ЦМО, участок ТО КУУН	18

Другим важным параметром при регистрации обращения является «Услуга».

Т.к. для некоторых подразделений, например «Системные администраторы», этот параметр является необязательным, то надо проверить данные на пустые значения и удалить их.

```
108 service_df['service'].replace('', np.nan, inplace=True)
109 service_df.dropna(subset=['service'], inplace=True)
```

Рисунок 4.2.2. Проверка и удаление пустых значений

Количество обращений, для которых определены услуги, составляет 7054. Количество услуг, оказанных пользователю, составляет 291. Т.к. одну и ту же услугу могут предоставлять несколько подразделений, то необходимо сгруппировать услуги по подразделениям.

```
111 countServ=len(Series(service_df['service'].values.ravel()).unique())
112 groupServGroup=service_df.pivot_table(['description'],['name','service'], aggfunc='count', fill_value = 0)
```

Рисунок 4.2.3. Количество обращений по группе исполнителей и услугам

Количество обращений по виду услуг представлены в таблице 4.2.2.

Таблица 4.2.2. Распределение количества обращений по группам исполнителей и оказываемым услугам

Группа исполнителей	*Услуга*	*Количество обращений*
Группа первой линии поддержки пользователей HP	Комплексное обслуживание мобильного компьютера	4
Группа первой линии поддержки пользователей HP	Комплексное обслуживание стационарного компьютера	212

Группа исполнителей	*Услуга*	*Количество обращений*
Группа сопровождения ВКС	Техническая поддержка совещаний ЦА ЮНГ в залах ВКС	26
Группа сопровождения ИС Инфраструктура	Поддержка пользователя ИС Инфраструктура на базе Maximo	42
Группа сопровождения ИС Инфраструктура	Управление данными. ПО Инфраструктура на базе Maximo	24
Сектор сопровождения систем технол. мониторинга	Поддержка БП ЦДС	10
Сектор сопровождения систем технол. мониторинга	Поддержка и развитие БП ЦДС-Консолидатор	25
Сектор сопровождения систем технол. мониторинга	Сопровождение БП ОПУС	3
Сектор сопровождения систем технол. мониторинга	Сопровождение БП ЦДС	40

Параметр «Категория» зависит только от описания обращения. Всего категорий по оказанному сервису 7. Количество обращений по категориям представлено в таблице 4.2.3.

Таблица 4.2.3. Распределение количества обращений категориям

Категория	*Количество обращений*
Запрос на изменение	73
Запрос на обслуживание	6370
Инцидент	191
Модернизация оборудования	108

Категория	*Количество обращений*
Ожидание от МОЛ	16
Постановка на обслуживание	667
Приобретение оборудования	23

Определив параметры, выполняется предобработка данных.

4.3. ОБУЧЕНИЕ КЛАССИФИКАТОРА

С помощью библиотеки NTLK, были подключены модули для загрузки списка стоп-слов, стемминга и регулярных выражений. Предобработку выполним для параметра «Описание».

```
47 stemmer = SnowballStemmer("russian")
48 tokenizer = RegexpTokenizer(r'\w+')
49 en_stop = stopwords.words('russian')

127 request_df['new_name'] =request_df['description']+' '+request_df['company'] +' '+ request_df['users']
128
129 request_df.new_name=request_df['new_name'].str.lower()
130
131 request_df.new_name=request_df['new_name'].apply(tokenizer.tokenize)#word_tokenize)
132 request_df.new_name=request_df['new_name'].apply(lambda x: [item for item in x if item not in en_stop])
133 request_df.new_name=request_df['new_name'].apply(lambda x: [stemmer.stem(i) for i in x])
```

Рисунок 4.3.1. Предобработка текста

После предобработки текста строится матрица терм-документ на основе словаря коллекции. Метод CountVectorizer() – производит подсчет терма в документе. Значение индекса слова в словаре связано с его частотой употребления во всем корпусе [15].

```
159 vectorizer = CountVectorizer()
160
161 X = vectorizer.fit_transform(x for x in texts2).A
162 vn=vectorizer.get_feature_names()
```

Рисунок 4.3.2. Построение матрицы употребления терм-документ

Пример терма с его индексом в словаре представлен в таблице 4.3.1.

Таблица 4.3.1. Словарь терминов

Индекс	*Слово*
2690	акта_
2691	актив
2692	активац
2694	активиров
…	…
4383	индикац
4384	инженер
….	…
8640	юганскнефтегаз
8641	югр

После построения матрицы употреблений терминов в документе, необходимо преобразовать ее в представление TF-IDF, чтобы определить вес терма. Преобразование матрицы в частотную матрицу TF-IDF представлено на рисунке 4.3.3.

```
113 vectorizer = CountVectorizer()
114
115 X = vectorizer.fit_transform(x for x in texts2).A
116
117 vn=vectorizer.get_feature_names()
118 dictionary=vectorizer.vocabulary_
119
120 print('TF-IDF')
121 tfidf_transformer = TfidfTransformer()
122 X_train_tfidf = tfidf_transformer.fit_transform(X).A
123 X_train_tfidf.shape
```

Рисунок 4.3.3. Определение веса термина

Описание модели классификации представлено на рисунке 4.3.4.

```
clf=MultinomialNB().fit(X_train_tfidf,valsgroup)

predicted = clf.predict(X_train_tfidf)

dfpr=pd.DataFrame({'des':request_df['description'],'name': request_df['name'],'pred':predicted})

print (np.mean(predicted == valsgroup))

metric=metrics.classification_report(valsgroup, predicted, target_names=UniqGroup)
```

Рисунок 4.3.4. Построение классификатора

Используя данные параметры для обучения, классификатор верно распределил 78% обращений по подразделениям. Пример полученных значений прогноза представлен в таблице 4.3.5.

Таблица 4.3.5. Прогноз системы

Описание	*Группа исполнителей*	*Прогноз системы*
Выдать оборудование – фотоаппарат.	Системные администраторы Мамонтовского региона	Системные администраторы Правдинского региона
“Замена существующего оборудования Стационарный ПК”	Системные администраторы Правдинского региона	Системные администраторы Правдинского региона
При распечатывании документов в принтере происходит шум (печка не разбавляет кра	Сервисный центр	Сервисный центр
Выдать оборудование – Фотоаппарат	Системные администраторы Правдинского региона	Системные администраторы Правдинского региона
Замена существующего оборудования – Стационарный ПК	Системные администраторы ПЗ Пионерная 5П	Сервисный центр
Заменить RM1-6303-000000 -20 шт. Тормозные площадки	Сервисный центр	Сервисный центр

Описание	*Группа исполнителей*	*Прогноз системы*
Замена аккумуляторных батарей на ИБП инв. 101831721	Сервисный центр	Сервисный центр
Подозрительная активность 10.227.215.23 10.227.215.25 РН-Юганскнефтегаз	ТО КСБ Майского и Мамонтовского р-на	Сервисный центр
Заменить	Сервисный центр	Сервисный центр
Выдать оборудование - Стационарный ПК	Системные администраторы Майского региона	Системные администраторы Правдинского региона
Выдать оборудование. Монитор	Системные администраторы Правдинского региона	Системные администраторы Правдинского региона
Выдать оборудование. Стационарный ПК	Системные администраторы Правдинского региона	Системные администраторы Правдинского региона
Добавить в классификатор OIS насос согласно приложения.	Сектор сопровожд. Систем нефтедобычи и разработки	Сектор сопровожд. Систем нефтедобычи и разработки
Настроить подключение к ЕИСИПиМ.	Сектор сопровожд. Систем нефтедобычи и разработки	Сектор сопровожд. Систем нефтедобычи и разработки

Выполнив оценку классификатора, можно определить, что классификатор не распознает подразделения, количество обращений по которым < 100.

Точность классификатора на данной выборке составляет 84%, то есть классификатор принял верное решение по 84% обращений в выборке, полнота – 73%.

Усредненное значение точности и полноты составляет 78%.

На необработанных данных классификатор верно распределяет 70% обращений.

Для более точного прогноза по определению подразделений были расширены параметры. Т.к. для каждого подразделения определены свои услуги, то был добавлен параметр «Услуга».

Оценка качества классификатора приведена на рисунке 4.3.5.

	precision	recall	f1-score	support
Группа первой линии поддержки пользователей НР	0.60	0.94	0.73	223
Группа первой линии поддержки пользователей ПР	0.00	0.00	0.00	1
Группа сопровождения ВКС	1.00	0.23	0.38	26
Группа сопровождения ИС Инфраструктура	1.00	0.58	0.73	66
Дежурные администраторы НО	0.00	0.00	0.00	6
Диспетчерская служба (ИТ, связь)	0.00	0.00	0.00	2
Зарплата и Кадры	1.00	0.20	0.33	56
ИС МТО	0.76	0.74	0.75	129
ИС Финансового Управления ЮНГ	0.00	0.00	0.00	41
ОИС нормативно-справочной информации	1.00	0.47	0.64	59
ОС АСУ ТП - Мамонтовский и Майский регион	0.74	0.52	0.61	194
ОС АСУ ТП - Приобский регион	1.00	0.18	0.31	61
ОС АСУ ТП - Юганский регион	0.00	0.00	0.00	57
ОС АСУ ТП - Пойковский регион	1.00	0.14	0.24	74
ОС и Налоги	1.00	0.03	0.05	39
Отдел АТС и линий связи	0.56	0.98	0.71	418
Отдел администрирования СУБД	1.00	0.18	0.31	33
Отдел главного энергетика	0.00	0.00	0.00	15
Отдел информационной безопасности	0.74	0.86	0.80	116
Отдел информационных систем ФХД	0.90	0.80	0.85	127
Отдел подготовки данных	1.00	0.39	0.56	89
Отдел системной инженерии	0.89	0.65	0.75	169
Отдел телекоммуникационных радиотехнологий	1.00	0.09	0.17	54
Поддержка ИС DIRECTUM	0.59	0.98	0.74	317
Проектная группа сопровождения ЭПОС	0.93	0.98	0.95	292
Сектор связи ИТ	0.00	0.00	0.00	3
Сектор сопровожд. систем нефтедобычи и разработки	0.68	0.90	0.78	161
Сектор сопровождения оперативных систем (ССОС)	0.95	0.77	0.85	137
Сектор сопровождения систем технол. мониторинга	0.82	0.63	0.71	102
Сервисный центр	0.41	0.97	0.57	374
Сервисный центр Правдинского региона	0.00	0.00	0.00	76
Сервисный центр Пыть-Яхского отделения	1.00	0.31	0.47	88
Системные администраторы DFS	0.76	0.95	0.85	139
Системные администраторы Культурного центра	0.00	0.00	0.00	31
Системные администраторы Майского региона	0.95	0.42	0.58	138
Системные администраторы Мамонтовского региона	1.00	0.17	0.28	109
Системные администраторы ПЗ Пионерная 5П	0.82	0.18	0.30	76
Системные администраторы Правдинского региона	0.50	0.91	0.64	257
Системные администраторы Приобского - Правый берег	1.00	0.07	0.13	88
Системные администраторы УДНГ ЮР, Юг. регион	0.89	0.40	0.55	77

Системные администраторы УДНГ ЮР, Юг. регион	0.89	0.40	0.55	77
Системные администраторы зд Жилая 20	1.00	0.05	0.10	60
Системные администраторы зд Нефтяников 1	0.00	0.00	0.00	1
Системные администраторы здания 15-13	0.00	0.00	0.00	22
Системные администраторы здания 5-16	0.00	0.00	0.00	16
Системные администраторы здания 7-49	1.00	0.19	0.33	67
Системные администраторы здания Л26	0.92	0.41	0.56	118
Системные администраторы здания Л26А	1.00	0.17	0.29	101
Системные администраторы здания НКИ	0.00	0.00	0.00	20
Системные администраторы здания УКС	1.00	0.11	0.20	115
Системные администраторы здания УМТО	1.00	0.12	0.22	66
Системные администраторы здания УНС	0.54	0.92	0.68	252
Системные администраторы здания УРИ	0.00	0.00	0.00	30
Системные администраторы здания ФЭЦ	0.20	0.09	0.12	49
Системный администраторы здания РН-Учет	0.85	0.87	0.86	161
Сопровождение БП "1С"	0.00	0.00	0.00	9
Сопровождение задач основного производства ПяО	0.89	0.35	0.51	93
ТО КСБ Майского и Мамонтовского р-на	1.00	0.14	0.25	64
ТО КСБ Пойковского р-на и лев.берега Приобского	1.00	0.02	0.04	50
ТО КСБ ЮР и прав.берега Приобского	0.69	0.72	0.71	68
ТО ОУУ Майского р-на и Мамонтово	0.00	0.00	0.00	18
ТО ОУУ Пойковского р-на	0.00	0.00	0.00	24
ТО ОУУ Юганского и Приобского р-н	0.00	0.00	0.00	6
Технологи Правдинского региона	1.00	0.09	0.17	64
УС Майского и Мамонтовского р-на	0.00	0.00	0.00	28
УС Пойковског региона и левый берег Приобского	0.00	0.00	0.00	11
УС Юганского р-на и правый берег Приобского	0.00	0.00	0.00	20
Удаленный мониторинг бурения	0.00	0.00	0.00	21
Участок монтажно-наладочных работ	0.00	0.00	0.00	2
Участок обслуживания КС	0.00	0.00	0.00	11
Участок по ТО комплексных систем безопасности	0.00	0.00	0.00	7
Учет капитальных вложений	0.94	0.43	0.59	114
Финансы и Отчетность	0.98	0.64	0.77	77
ЦАП-1 (Юганский регион)	0.81	0.95	0.87	225
ЦАП-2 (Приобский регион)	1.00	0.02	0.05	43
ЦАП-3 (Майский регион)	0.52	0.99	0.68	348
ЦАП-4 (Мамонтово)	0.67	0.99	0.80	326
ЦАП-5	0.66	0.97	0.78	202
ЦМО, участок ТО КУУН	0.00	0.00	0.00	18
avg / total	0.84	0.73	0.78	7448

Рисунок 4.3.5. Оценка качества классификатора

При достаточно хорошо обученном классификаторе можно оценить качество классификатора на тестовой выборке.

Количество строк в тестовой выборке составляет 4965. Количество классов для классификации по параметрам представлено в таблице 4.3.6.

Таблица 4.3.6. Количество классов для классификации в тестовой выборке

Параметр	*Количество*
Подразделение	76
Услуга	264
Категория	7

Точность классификации обращений пользователей по подразделениям на тестовой выборке составляет 80%.

```
Xt = vectorizer.transform(x for x in texts2t).A

UniqGroupt=request_dft['name'].sort_values().unique()

print('TF-IDF')
X_train_tfidft = tfidf_transformer.transform(Xt).A

predictedt = clf.predict(X_train_tfidft)

dfprt=pd.DataFrame({'des':request_dft['description'],'name': request_dft['name'],'pred':predictedt})

print (np.mean(predictedt == request_dft['name']))

metrict=metrics.classification_report(request_dft['name'], predictedt, target_names=UniqGroupt)
print(metric)
```

Рисунок 4.3.6. Проверка работы классификатора на тестовой выборке

	precision	recall	f1-score	support
Группа первой линии поддержки пользователей НР	0.50	0.95	0.65	106
Группа сопровождения ВКС	1.00	0.85	0.92	26
Группа сопровождения ИС Инфраструктура	1.00	0.95	0.97	38
Дежурные администраторы НО	0.00	0.00	0.00	3
Диспетчерская служба (ИТ, связь)	1.00	0.12	0.21	22
Зарплата и Кадры	1.00	0.75	0.86	44
ИС МТО	0.80	0.98	0.88	104
ИС Финансового Управления ЮНГ	0.00	0.00	0.00	23
ОИС нормативно-справочной информации	1.00	0.34	0.51	29
ОС АСУ ТП - Мамонтовский и Майский регион	0.60	0.90	0.72	123
ОС АСУ ТП - Приобский регион	1.00	0.11	0.20	36
ОС АСУ ТП - Юганский регион	1.00	0.27	0.43	44
ОС АСУ ТП - Пойковский регион	1.00	0.03	0.05	39
ОС и Налоги	1.00	0.59	0.74	29
Отдел АТС и линий связи	0.74	1.00	0.85	318
Отдел администрирования СУБД	1.00	0.16	0.28	25
Отдел главного энергетика	0.00	0.00	0.00	13
Отдел информационной безопасности	0.95	0.95	0.95	73
Отдел информационных систем ФХД	0.95	0.77	0.85	78
Отдел подготовки данных	1.00	0.71	0.83	69
Отдел системной инженерии	0.86	0.74	0.80	120
Отдел телекоммуникационных радиотехнологий	1.00	0.03	0.06	32
Поддержка ИС DIRECTUM	0.72	1.00	0.83	234
Проектная группа сопровождения ЭПОС	0.97	1.00	0.98	200
Сектор сопровожд. систем нефтедобычи и разработки	0.70	0.97	0.81	118
Сектор сопровождения оперативных систем (ССОС)	0.90	0.82	0.86	107
Сектор сопровождения систем технол. мониторинга	0.79	0.79	0.79	72
Сервисный центр	0.53	0.98	0.69	262
Сервисный центр Правдинского региона	0.00	0.00	0.00	59
Сервисный центр Пыть-Яхского отделения	0.96	0.44	0.61	54
Системные администраторы DFS	0.89	1.00	0.94	102
Системные администраторы Культурного центра	0.00	0.00	0.00	15
Системные администраторы Майского региона	0.96	0.47	0.63	96
Системные администраторы Мамонтовского региона	1.00	0.27	0.43	59
Системные администраторы ПЗ Пионерная 5П	1.00	0.36	0.53	55
Системные администраторы Правдинского региона	0.60	0.91	0.72	204
Системные администраторы Приобского - Правый берег	1.00	0.15	0.27	65
Системные администраторы УДНГ ЮР, Юг. регион	0.88	0.46	0.60	61
Системные администраторы зд Жилая 20	1.00	0.31	0.47	55
Системные администраторы зд Нефтяников 1	0.00	0.00	0.00	1
Системные администраторы здания 15-13	0.00	0.00	0.00	11

Системные администраторы здания 5-16	0.00	0.00	0.00	11
Системные администраторы здания 7-49	1.00	0.30	0.47	46
Системные администраторы здания Л26	0.98	0.69	0.81	72
Системные администраторы здания Л26А	0.97	0.53	0.68	55
Системные администраторы здания НКИ	0.00	0.00	0.00	9
Системные администраторы здания УКС	1.00	0.31	0.47	62
Системные администраторы здания УМТО	1.00	0.05	0.09	44
Системные администраторы здания УНС	0.52	1.00	0.68	161
Системные администраторы здания УРИ	0.00	0.00	0.00	16
Системные администраторы здания ФЭЦ	0.00	0.00	0.00	35
Системный администраторы здания РН-Учет	0.92	0.97	0.94	79
Сопровождение БП "1С"	0.00	0.00	0.00	7
Сопровождение задач основного производства ПяО	0.91	0.38	0.53	53
ТО КСБ Майского и Мамонтовского р-на	0.53	0.21	0.30	43
ТО КСБ Пойковского р-на и лев.берега Приобского	1.00	0.25	0.40	28
ТО КСБ ЮР и прав.берега Приобского	0.69	0.53	0.60	51
ТО ОУУ Майского р-на и Мамонтово	0.00	0.00	0.00	12
ТО ОУУ Пойковского р-на	0.10	0.06	0.07	20
ТО ОУУ Юганского и Приобского р-н	0.00	0.00	0.00	3
Технологи Правдинского региона	1.00	0.25	0.39	53
УС Майского и Мамонтовского р-на	0.00	0.00	0.00	14
УС Пойковског региона и левый берег Приобского	0.00	0.00	0.00	8
УС Юганского р-на и правый берег Приобского	0.00	0.00	0.00	15
Удаленный мониторинг бурения	1.00	0.06	0.11	18
Участок монтажно-наладочных работ	0.00	0.00	0.00	2
Участок обслуживания КС	0.00	0.00	0.00	11
Участок по ТО комплексных систем безопасности	0.00	0.00	0.00	7
Учет капитальных вложений	1.00	0.77	0.87	75
Финансы и Отчетность	0.86	0.90	0.88	61
ЦАП-1 (Юганский регион)	0.84	0.92	0.88	140
ЦАП-2 (Приобский регион)	1.00	0.04	0.08	23
ЦАП-3 (Майский регион)	0.61	1.00	0.76	201
ЦАП-4 (Мамонтово)	0.74	0.97	0.84	195
ЦАП-5	0.74	0.90	0.81	118
ЦЮО, участок ТО КУУН	0.00	0.00	0.00	17
avg / total	0.80	0.69	0.74	4965

Рисунок 4.3.7. Оценка качества классификатора

Для более подробного анализа по распределению обращений, используя модуль для оценки качества классификации – sklearn.metrics, была построена матрица неточностей (Cij).

Элементы матрицы Cij – это количество объектов, относящихся реально к классу i, но по прогнозной модели отнесенных к классу j [15,17].

На рисунке 4.3.8. приведена матрица неточностей для подразделений, по которым количество обращений > 100.

Проанализировав данные, можно сделать вывод, что большинство документов классификатор распределяет однозначно верно. Из 120 обращений по подразделению «Отдел системной инженерии» классификатор 91 обращение определил верно, а 29 обращений распределил на такие подразделения как: «Группа первой линии поддержки пользователей HP», «Отдел информационных систем ФХД», «Поддержка ИС DIRECTUM», «Системные администраторы DFS» и «Системные администраторы Правдинского региона».

Матрица неточностей для подразделений, по которым количество обращений < 100, представлена в Приложении 4.

После распределения обращения по подразделениям необходимо назначить данному обращению категорию.

Выполнив обучение и сравнив данные, которые выдала система и данные в выборке, можно сделать вывод, что данные в выборке не совсем достоверны.

Например, обращение «поставить на обслуживание» может принимать категорию «Постановка на обслуживание» и «Запрос на обслуживание», поскольку количество данных обращений больше у категории «Постановка на обслуживание», то достовернее относить такие обращения к данной категории. Система распределила верно 92% обращений по категориям.

```
clfcat=MultinomialNB().fit(X_train_tfidfcat,category_df['category'])

UniqGroupcatt=category_dft['category'].sort_values().unique()

#print('TF-IDF')
X_train_tfidfcatt = tfidf_transformer.transform(Xcatt).A

predictedcatt = clfcat.predict(X_train_tfidfcatt)
dfprcatt=pd.DataFrame({'des':category_dft['description'],'cat': category_dft['category'],'pred':predictedcatt})
print (np.mean(predictedcatt == category_dft['category']))

metriccatt=metrics.classification_report(category_dft['category'], predictedcatt, target_names=UniqGroupcatt)
print(metriccatt)

confus_catt=metrics.confusion_matrix(category_dft['category'], predictedcatt, labels=None, sample_weight=None)
print(confus_catt)
```

Рисунок 4.3.9. Проверка классификатора

		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
0	Группа первой линии поддержки пользователей HP	101	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0
1	ИС МТО	0	102	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	ОС АСУ ТП - Мамонтовский и Майский регион	0	0	111	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	4	0
3	Отдел АТС и линий связи	0	0	0	317	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
4	Отдел информационной безопасности	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	Отдел подготовки данных	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	Отдел системной инженерии	8	0	0	0	2	0	91	3	0	1	0	6	4	4	0	0	1	0	0	0	0
7	Поддержка ИС DIRECTUM	0	0	0	1	0	1	0	232	0	0	0	0	0	0	0	0	0	0	0	0	0
8	Проектная группа сопровождения ЭПОС	0	0	0	0	0	0	0	0	200	0	0	0	0	0	0	0	0	0	0	0	0
9	Сектор сопровожд. систем нефтедобычи и разработки	0	0	0	1	0	0	0	2	0	114	0	0	1	0	0	0	0	0	0	0	0
10	Сектор сопровождения оперативных систем (ССОС)	0	0	0	3	0	1	0	5	0	0	83	0	0	0	0	0	0	1	0	0	0
11	Сервисный центр	0	0	0	0	0	0	0	0	0	0	0	258	0	1	3	0	0	0	0	0	0
12	Системные администраторы DFS	0	0	0	0	0	0	0	0	0	0	0	0	102	0	0	0	0	0	0	0	0
13	Системные администраторы Правдинского региона	0	0	0	3	0	0	0	1	0	0	1	7	1	187	0	3	0	0	0	0	1
14	Системные администраторы здания УКС	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	Системные администраторы здания УНС	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	161	0	0	0	0	0
16	Системный администраторы здания РН-Учет	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	ЦАП-1 (Юганский регион)	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	129	7	1	2
18	ЦАП-3 (Майский регион)	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	200	0	0
19	ЦАП-4 (Мамонтово)	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	190	0
20	ЦАП-5	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	1	9	1	106

Рисунок 4.3.8. Матрица неточностей

	precision	recall	f1-score	support
Запрос на изменение	1.00	0.59	0.74	44
Запрос на обслуживание	0.93	1.00	0.96	4196
Инцидент	0.91	0.43	0.58	101
Модернизация оборудования	0.00	0.00	0.00	85
Ожидание от МОЛ	0.00	0.00	0.00	7
Постановка на обслуживание	0.99	0.77	0.87	510
Приобретение оборудования	0.00	0.00	0.00	22
avg / total	0.92	0.94	0.92	4965

Рисунок 4.3.10. Оценка качества классификатора

Для подробного анализа распределения обращений по категориям была построена матрица неточностей.

```
cm =metrics.confusion_matrix(category_dft['category'], predictedcatt)
classes = UniqGroupcatt
cmap = plt.cm.Blues
plt.imshow(cm, interpolation='nearest',cmap=cmap)
plt.title('Confusion matrix')
plt.colorbar()
tick_marks = np.arange(len(classes))
plt.xticks(tick_marks, classes, rotation=45)
plt.yticks(tick_marks, classes)

fmt = 'd'
thresh = cm.max() / 2.
for i, j in itertools.product(range(cm.shape[0]),range(cm.shape[1])):
    plt.text(j, i, format(cm[i, j], fmt),horizontalalignment="center", color="white" if cm[i, j] > thresh else "black")
plt.tight_layout()
plt.ylabel('True label')
plt.xlabel('Predicted label')

plt.savefig("confusion_matrix.png")
```

Рисунок 4.3.11. Построение матрицы ошибок

Проанализировав матрицу неточностей, можно выявить, какие категории плохо различает классификатор.

На рисунке 4.3.12 можно определить, что классификатор не распознает категории, обращений по которым достаточно мало.

Категорию «Инцидент» в большей степени относит к категории «Запрос на обслуживание»

Категорию «Запрос на изменение» в 60% определяет верно, а 40% относит к категории «Запрос на обслуживание»

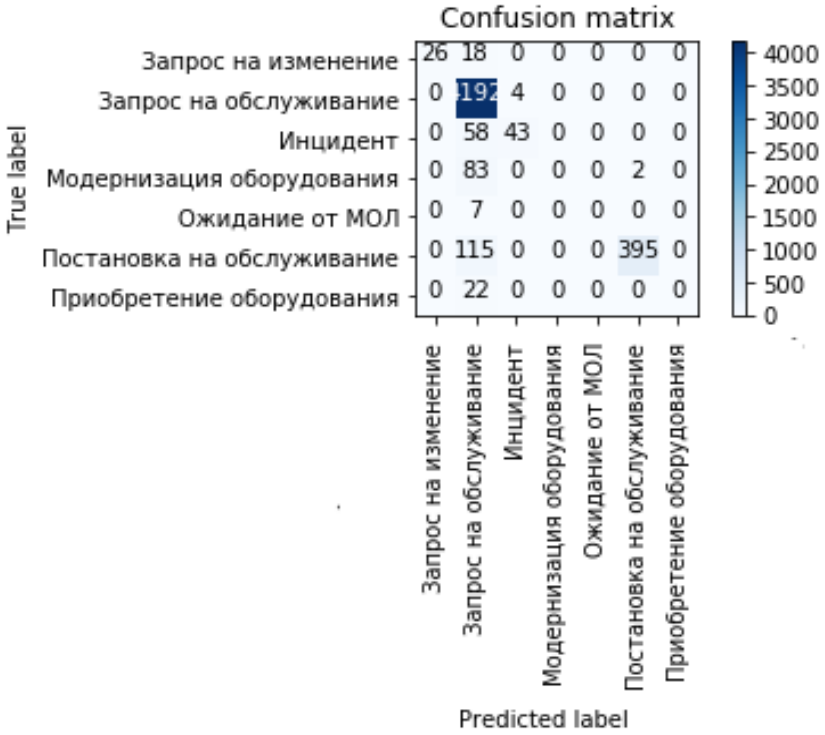


Рисунок 4.3.12. Матрица неточностей

ЗАКЛЮЧЕНИЕ

Подведя итоги в сравнении ручного и автоматизированного регистрирования обращений, можно выделить большой перечень плюсов, которые можно достичь при минимальных затратах, а именно: снижение количества времени, повышение производительности, уменьшение влияния человеческого фактора, приводящего к неверным результатам, возможность сконцентрироваться на остальных не менее важных задачах.

В рамках проекта выполнены задачи:

- Рассмотрены и проанализированы методы классификации текста.
- Выбран язык программирования и инструментарий для разработки системы.
- Подготовлена обучающая выборка, на основе истории журнала заявок.
- Реализована система для обработки обращений пользователей и выдачи результата в качестве ответственного подразделения.

СПИСОК ЛИТЕРАТУРЫ

1. Агеев М. С., Добров Б. В., Лукашевич Н. В. Автоматическая рубрикация текстов: методы и проблемы: Учёные записки Казанского государственного университета. Серия Физико-математические науки, 2008. — Т. 150, № 4. — С. 25–40.
2. Батура Т.В. Методы автоматической классификации текстов: Программные продукты и системы, 2017. — Т. 30. № 1. — С. 85–99
3. Большакова Е.И., Воронцов К.В., Ефремова Н.Э., Клышинский Э.С., Лукашевич Н.В., Сапин А.С: Автоматическая обработка текстов на естественном языке и анализ данных: учеб. Пособие — М.: Изд-во НИУ ВШЭ, 2017. — С. 269.
4. Бурлаева Е.И., Обзор методов классификации текстовых документов на основе подхода машинного обучения: Программная инженерия: Изд-во Новые технологии, 2017. — 328–336 с.
5. Плас Дж. Вандер. Python для сложных задач: наука о данных и машинное обучение. — СПБ.: Питер, 2018. — 576 с.
6. Вежневец Владимир. Оценка качества работы классификаторов. Компьютерная графика и мультимедиа. Выпуск №4(1),2006.
7. Дремина А.К. Определение полезности признаков в задаче классификации коротких текстовых сообщений: Молодежный научно-технический вестник: Изд-во ФГБОУ ВПО «МГТУ им. Н.Э.Баумана», 2013 №10 — с. http://sntbul.bmstu.ru/doc/617146.html
8. А.С. Епрев Автоматическая классификация текстовых документов: Математические структуры и моделирование 2010, вып. 21, с 65-81.
9. Павел Жук, Мария Давыдова, Полуавтоматическое извлечение часто задаваемых вопросов из обращений в службу поддержки: Second Conference on Software Engineering and Information Management, 2017. — 6 -12 с.

10. Андреас Мюллер, Сара Гвидо: Введение в машинное обучение с помощью Python. Руководство для специалистов по работе с данными, 2017— 480 с.
11. Себастьян Рашка: Python и машинное обучение, 2017 — 418 с.
12. Документация для Фреймворка Flask URL: http://flask-russian-docs.readthedocs.io/ru/latest/tutorial/ (Дата обращения: 10.05.2018)
13. Анализ текстов: от наивного Байеса до тематического моделирования, URL: https://logic.pdmi.ras.ru/~sergey/teaching/mlbeeline16/N16_BeelineTextMining.pdf. (Дата обращения: 05.03.2018)
14. Видеокурсы Яндекс по машинному обучению. URL: https://yandexdataschool.ru/edu-process/courses/machine-learning. (Дата обращения: 10.03.2018).
15. Документация Scikit-learn. URL: http://scikit-learn.org/stable/index.html (Дата обращения: 15.06.2018).
16. Наивный байесовский классификатор. URL: http://scikit-learn.org/stable/modules/naive_bayes.html (Дата обращения 10.06.2018).
17. Оценка качества классификатора. URL: http://bazhenov.me/blog/2012/07/21/classification-performance-evaluation.html (Дата обращения 10.06.2018).
18. Многоклассовая классификация текста. Метрики качества. URL: https://www.coursera.org/lecture/vvedenie-mashinnoe-obuchenie/mnoghoklassovaia-klassifikatsiia-P9Zun (Дата обращения 15.06.2018).

ПРИЛОЖЕНИЕ 1.

Руководство пользователя

После загрузки web-страницы, пользователь может оформить заявку для решения проблемы.

Для этого пользователю необходимо в текстовом окне ввести описание своей проблемы. После чего, по нажатию на кнопку «Ок» выдается описание проблемы и рекомендация для назначения ответственного подразделения. (см. рисунок.1.1)

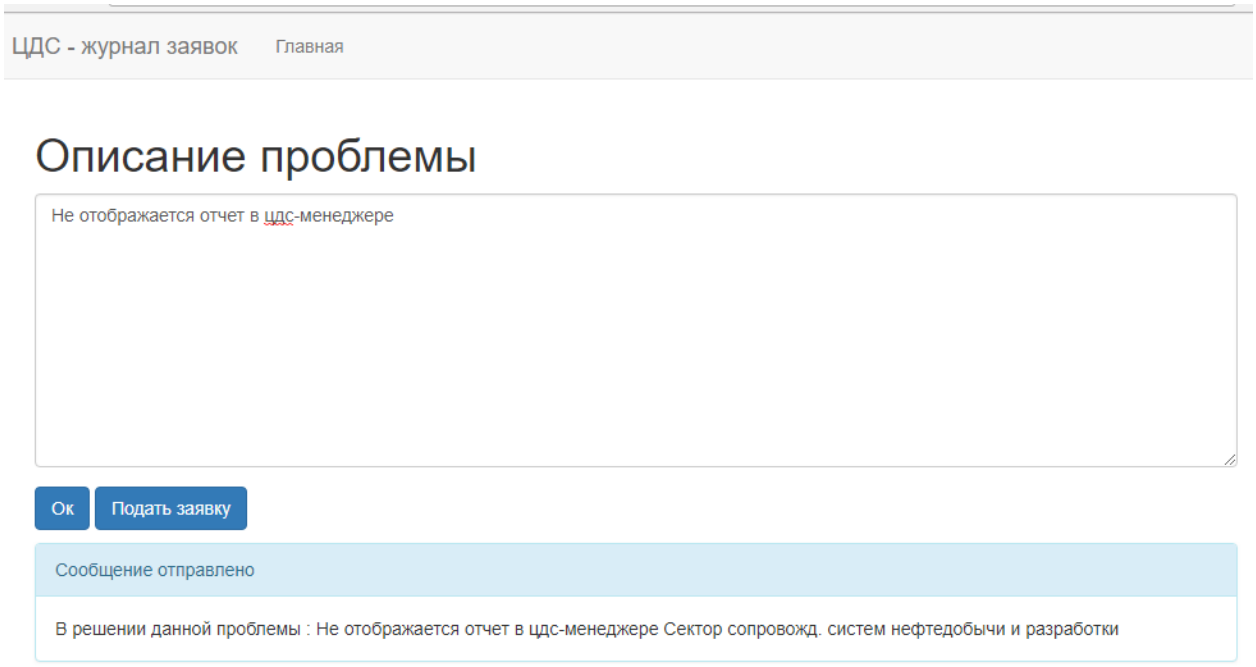


Рисунок 1.1.Форма ввода описания заявки и вывода рекомендации

В случае, если пользователь согласен с результатом, результат сохранится в базе данных и заявка отобразится на сайте ЦДС-Журнал заявок. (см. Рисунок 1.2.)

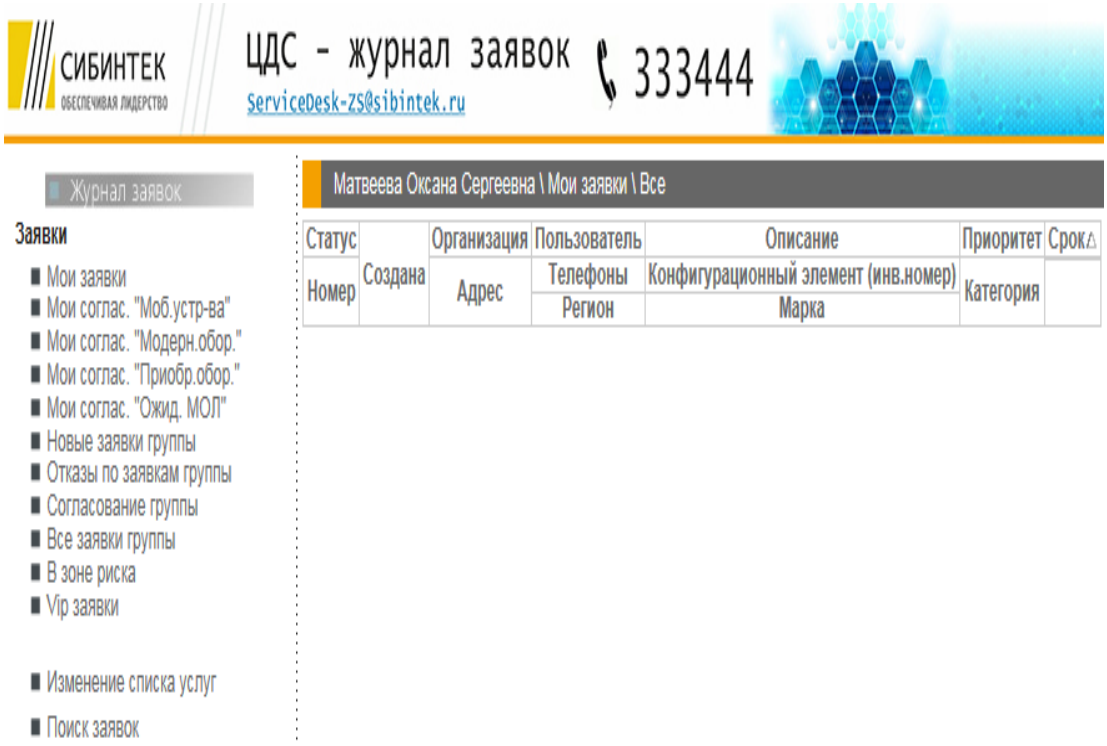


Рисунок 1.2. Система ЦДС-Журнал заявок

ПРИЛОЖЕНИЕ 2.

Описание методов для обработки нового обращения в скрипте Classification. py

Предобработка текста, введенного пользователем.

```
def analyzer_text (text):
    new_s=''
    textanalize=[]
    text=text.lower()
    tokens=tokenizer.tokenize(text)#word_tokenize)
    stopped_tokens = [i for i in tokens if not i in en_stop]
    stemmed_tokens = [stemmer.stem(i) for i in stopped_tokens]
    new_s=' '.join([morph.parse(i)[0].normal_form for i in stemmed_tokens])
    textanalize.append(new_s)
    new_s=''
    return textanalize
```

Определение подразделения по обращению пользователя.

```
def get_class(docs_new):
    text= analyzer_text (docs_new)
    with open('model.pkl', 'rb') as fin:
        tfidf_transformer, clf = pickle.load(fin)
    X_new_counts = vectorizer.transform(text)
    X_new_tfidf = tfidf_transformer.transform(X_new_counts)
    predictednew = clf.predict(X_new_tfidf)
    return predictednew
```

ПРИЛОЖЕНИЕ 3.

Скрипт main_app.py

Вывод на форму рекомендации по назначению группы исполнителей.

```
from flask import Flask, render_template, request
from Classification import get_class
app = Flask(__name__)
@app.route('/')
def hello():
    if request.args:
        text = request.args['text']
        theme = get_ class (text)
        return render_template('index.html', result='В решении данной проблемы :
'+'\n'+ text + '\n' + theme)
    return render_template('index.html')
if __name__ == '__main__':
    app.run(debug=True)
```

ПРИЛОЖЕНИЕ 4.

Матрица неточностей для подразделений, по которым количество обращений < 100

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58
Группа первой линии поддержки пользователей НР	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Группа сопровождения ВКС	1	0	23	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Группа сопровождения ИС Инфраструктура	2	0	0	36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Диспетчерская служба (ИТ, связь)	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Зарплата и Кадры	4	0	0	0	0	35	6	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ИС МТО	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ОИС нормативно-справочной информации	6	1	0	0	0	0	0	10	0	0	0	0	0	3	0	0	0	0	0	0	4	0	0	0	0	8	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0
ОС АСУ ТП - Мамонтовский и Майский регион	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ОС АСУ ТП - Приобский регион	8	0	0	0	0	0	0	0	24	5	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	1	1	0	0	0
ОС АСУ ТП - Юганский регион	9	0	0	0	0	0	0	0	15	0	14	0	0	2	0	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	7	0	3	0	0	0
ОС АСУ ТП - Пойковский регион	10	0	0	0	0	0	0	0	20	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	2	12	0
ОС и Налоги	11	0	0	0	0	0	0	0	0	0	0	17	3	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	0	0	0	0	0
Отдел АТС и линий связи	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Отдел администрирования СУБД	13	0	0	0	0	0	0	0	0	0	0	0	0	6	0	0	0	8	0	0	0	1	0	2	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	3	0	0	0	0
Отдел информационной безопасности	14	2	0	0	0	0	0	0	0	0	0	0	0	0	69	0	0	0	0	1	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Отдел информационных систем ФХД	15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	63	0	0	0	10	0	2	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Отдел подготовки данных	16	0	0	0	0	0	0	0	0	0	0	0	4	0	0	0	51	0	0	3	0	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Отдел системной инженерии	17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Отдел телекоммуникационных радиотехнологий	18	0	0	0	0	0	0	0	0	0	0	0	27	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	0	0	0
Поддержка ИС DIRECTUM	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Проектная группа сопровождения ЭПОС	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Сектор сопровожд. систем нефтедобычи и разработки	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Сектор сопровождения оперативных систем (СООС)	22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Сектор сопровождения систем технол. мониторинга	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	7	0	57	1	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0
Сервисный центр	24	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Сервисный центр Правдинского региона	25	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	50	5	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Сервисный центр Пыть-Яхского отделения	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	27	0	26	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Системные администраторы DFS	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Системные администраторы Культурного центра	28	8	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Системные администраторы Майского региона	29	7	0	0	0	0	0	0	1	0	0	0	0	2	0	0	0	0	1	0	1	1	0	0	4	0	1	0	0	50	0	0	8	0	0	0	0	0	0	0	0	0	15	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0
Системные администраторы Мамонтовского региона	30	5	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	3	0	2	0	0	0	0	3	0	0	1	0	1	16	0	15	0	0	0	0	0	0	0	0	0	10	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0
Системные администраторы ПЗ Пионерная 5П	31	10	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	1	1	0	0	0	9	0	0	0	0	0	0	0	22	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

		0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58
Системные администраторы Правдинского региона	32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Системные администраторы Приобского - Правый берег	33	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	6	0	0	1	0	0	0	0	35	11	0	0	0	0	0	0	0	0	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Системные администраторы УДНГ ЮР, Юг. регион	34	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	5	0	0	0	0	0	0	0	13	0	28	0	0	0	0	0	0	0	11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Системные администраторы зд Жилая,20	35	8	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	9	0	0	0	0	0	0	0	2	0	0	26	0	0	0	0	0	0	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Системные администраторы здания 7-49	36	10	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	2	0	0	0	0	5	0	0	0	0	0	0	0	1	0	0	0	18	0	0	0	0	0	8	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0
Системные администраторы здания Л26	37	6	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	6	0	0	0	0	0	3	0	0	0	0	0	0	0	1	0	0	0	0	51	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Системные администраторы здания Л26А	38	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	8	0	0	0	0	0	0	0	1	0	0	0	0	1	31	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Системные администраторы здания НКИ	39	1	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Системные администраторы здания УКС	40	2	0	0	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	1	0	1	0	0	3	0	0	0	0	0	0	0	19	0	28	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0
Системные администраторы здания УМТО	41	18	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	14	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Системные администраторы здания УНС	42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Системный администраторы здания РН-Учет	43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Сопровождение задач основного производства ПяО	44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	1	2	4	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	20	0	0	0	0	0	0	0	3	0	12	3	0	0	0
ТО КСБ Майского и Мамонтовского р-на	45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ТО КСБ Пойковского р-на и лев.берега Приобского	46	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	7	4	0	0	0	0	0	0	3	2	8	0
ТО КСБ ЮР и прав.берега Приобского	47	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	6	0	27	0	0	0	0	0	3	0	2	4	0	0
ТО ОУУ Пойковского р-на	48	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	8	3	2	0	0
Технологи Правдинского региона	49	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	4	0	8	5	7	0	0	0	0	0	0	0	0	7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	16	0	0	0	0	0	1	0	4	0	0
УС Майского и Мамонтовского р-на	50	0	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	7	0	0	0	0
Удаленный мониторинг бурения	51	0	0	0	0	0	0	0	0	0	0	0	5	0	0	0	0	0	0	8	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	3	0	0	0	0	0
Учет капитальных вложений	52	0	0	0	0	0	0	0	0	0	0	0	5	0	0	2	0	0	0	4	2	2	0	0	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	58	0	0	0	0	0	0	0	0
Финансы и Отчетность	53	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	0	0	0	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	55	0	0	0	0	0	0	0
ЦАИ I-1 (Юганскии регион)	54	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ЦАП-2 (Приобский регион)	55	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	1	10	7	4	0	0	0
ЦАП-3 (Майский регион)	56	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ЦАП-4 (Мамонтово)	57	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
ЦАП-5	58	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0