

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ  
РОССИЙСКОЙ ФЕДЕРАЦИИ  
Федеральное государственное автономное образовательное учреждение  
высшего образования  
«ТЮМЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»  
ИНСТИТУТ МАТЕМАТИКИ И КОМПЬЮТЕРНЫХ НАУК  
Кафедра программной и системной инженерии

РЕКОМЕНДОВАНО К ЗАЩИТЕ В ГЭК

Заведующий кафедрой

*Д.т.н., профессор*

*А.Г. Ивашко*

2021 г.

**ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА**

магистерская диссертация

**СИСТЕМА МОНИТОРИНГА И АНАЛИЗА АКТИВНОСТИ  
ПОЛЬЗОВАТЕЛЯ НА ПЕРСОНАЛЬНОМ КОМПЬЮТЕРЕ**

09.04.03 Прикладная информатика

Магистерская программа «Информационные системы анализа данных»

Выполнила работу  
Студентка 2 курса  
очной формы обучения

Ульрих  
Маргарита  
Андреевна

Научный руководитель  
к.филол.н

Бидуля  
Юлия  
Владимировна

Рецензент  
к.ф.-м.н., доцент

Ступников  
Андрей  
Анатольевич

г. Тюмень, 2021

## РЕФЕРАТ

Отчет 52 с., 50 рис., 3 табл., 5 источн.

ВЕБ-ПРИЛОЖЕНИЕ, МОНИТОРИНГ, ПЕРСОНАЛ, МАШИННОЕ  
ОБУЧЕНИЕ, ВЕБ-СТРАНИЦА, РАБОЧЕЕ ВРЕМЯ, СТАТИСТИКА,  
КОНТРОЛЬ, КЛАССИФИКАТОР

Объектом работы является система для мониторинга деятельности за персональным компьютером.

Цель работы – разработка программного обеспечения для мониторинга деятельности за персональным компьютером.

Результаты работы: ПО, реализующее функции, необходимые для достижения поставленной цели.

## СОДЕРЖАНИЕ

РЕФЕРАТ	1
ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ	3
ВВЕДЕНИЕ	4
ГЛАВА 1. ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ МОНИТОРИНГА СОТРУДНИКОВ	6
1.1. АНАЛИЗ ПРОБЛЕМЫ НЕЭФФЕКТИВНОСТИ РАБОТЫ ПЕРСОНАЛА	6
1.2. ПРЕИМУЩЕСТВА ВНЕДРЕНИЯ ПО ДЛЯ МОНИТОРИНГА СОТРУДНИКОВ	11
1.3. ОБЗОР СУЩЕСТВУЮЩИХ СИСТЕМ	12
ГЛАВА 2. РАЗРАБОТКА ТРЕБОВАНИЙ И ПРОЕКТИРОВАНИЕ	15
2.1 ВЫЯВЛЕНИЕ ТРЕБОВАНИЙ	15
2.2. ПРОЕКТИРОВАНИЕ АРХИТЕКТУРЫ	18
2.3. ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ	19
ГЛАВА 3. КОНСТРУИРОВАНИЕ СИСТЕМЫ	22
3.1. ИНСТРУМЕНТЫ РАЗРАБОТКИ	22
3.2. РЕШЕНИЯ ПО РЕАЛИЗАЦИИ	22
3.3. МАШИННОЕ ОБУЧЕНИЕ ДЛЯ ОПРЕДЕЛЕНИЯ КАТЕГОРИИ ВЕБ-СТРАНИЦЫ	24
3.4. РАЗРАБОТКА ПОЛЬЗОВАТЕЛЬСКОГО ИНТЕРФЕЙСА	42
ЗАКЛЮЧЕНИЕ	49
СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ	51

## ПЕРЕЧЕНЬ СОКРАЩЕНИЙ И ОБОЗНАЧЕНИЙ

В настоящем отчете применяются следующие сокращения и обозначения.

БД	база данных
ОС	операционная система
ПК	персональный компьютер
СУБД	система управления базами данных
DFD	диаграммы потоков данных (Data Flow Diagrams)
ПО	программное обеспечение
UML	унифицированный язык моделирования (Unified Modeling Language)

## ВВЕДЕНИЕ

Исследования показывают, что большинство людей уделяет менее 40% рабочего времени непосредственно работе. Остальное время они занимаются своими делами, листая социальные сети, читая новости или играя в игры.

Такой неэффективный расход времени приводит к потере денег компаниями, и поэтому руководители стремятся различными способами контролировать сотрудников и следить, как и насколько эффективно они используют свое рабочее время

Один из способов контроля – специальное ПО, которое стало особенно актуально после пандемии Covid-19 и перехода на дистанционную работу. Учет рабочего времени удаленных сотрудников стал настоящей проблемой для руководства компаний, решить которую помогли программы для мониторинга удаленных сотрудников.

Функционал таких систем разнообразен, в частности, одной из главных функций является отслеживание активности сотрудников в сети. Однако, в ходе анализа существующих систем было выявлено, что их подход к мониторингу посещаемых веб-ресурсов не лишен недостатков.

На основании этого были сформулированы предложения по улучшению, позволяющие повысить качество и эффективность анализа. Таким образом, целью данной работы стала разработка программного обеспечения для мониторинга и анализа действий пользователей, в рамках которого были бы реализованы все предложенные улучшения.

Для успешной подготовки и защиты выпускной квалификационной работы автором ВКР использовались средства и методы физической культуры и спорта с целью поддержания должного уровня физической подготовленности, обеспечивающую высокую умственную и физической работоспособность. В режим рабочего дня включались различные формы

организации занятий физической культурой (физкультпаузы, физкультминутки, занятия бегом) с целью профилактики утомления, появления хронических заболеваний и нормализации деятельности различных систем организма.

В рамках подготовки к защите выпускной квалификационной работы автором созданы и поддерживались безопасные условия жизнедеятельности, учитывающие возможность возникновения чрезвычайных ситуаций.

# ГЛАВА 1. ПРОГРАММНОЕ ОБЕСПЕЧЕНИЕ ДЛЯ МОНИТОРИНГА СОТРУДНИКОВ

## 1.1. АНАЛИЗ ПРОБЛЕМЫ НЕЭФФЕКТИВНОСТИ РАБОТЫ ПЕРСОНАЛА

Британские исследователи опросили почти две тысячи офисных работников и выяснили, что все они продуктивно трудятся не более трех часов за рабочий день [1].

В первую очередь всем респондентам был задан вопрос: "Считаете ли вы себя продуктивным на протяжении всего рабочего дня?" 79% опрошенных ответили отрицательно. Затем исследователи попросили назвать время, в течение которого респонденты работают по-настоящему продуктивно, и получили среднее число — 2 часа 53 минуты.

Исходя из полученных данных, в среднем около 65 минут рабочего дня уходит на чтение новостей, 44 минуты на проверку социальных сетей, 40 минут на разговоры с коллегами (рисунок 1).

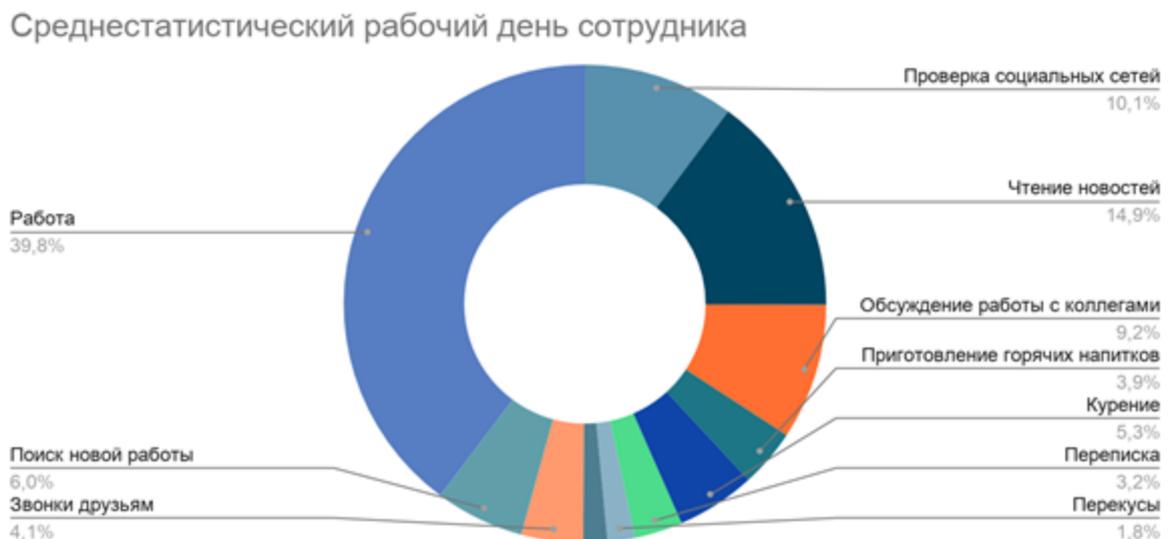


Рис.1. Среднестатистический рабочий день сотрудника

Источник: составлено автором на основе [1]

Среди пользователей Хабр был проведен опрос “Какой процент рабочего времени вы реально работаете, а не бездельничаете?” [2]. В опросе приняли участие 4156 пользователей. Как видно из результатов (рисунок 2), более 35% опрошенных выбрали вариант - 25%, а почти 5% проголосовали за “другое”, что скорее всего означает <25%. Таким образом, по результатам данного исследования чуть меньше половины продуктивно используют только 25% рабочего времени.

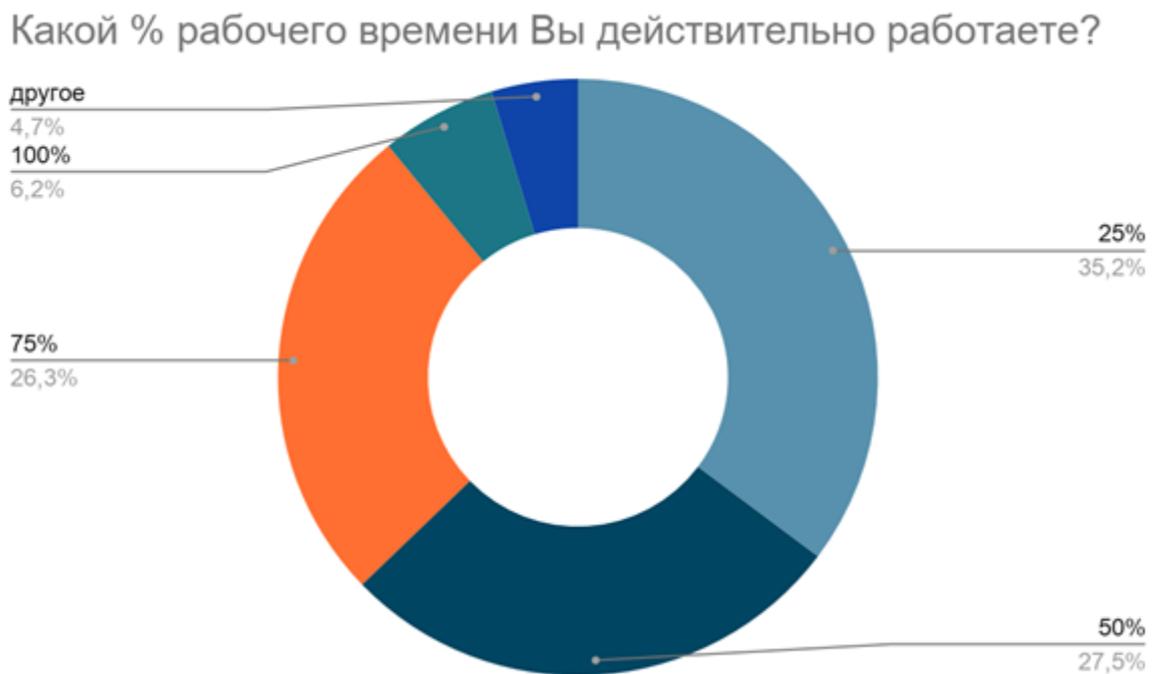


Рис.2. Результаты опроса пользователей Хабр

Источник: составлено автором на основе [2]

Согласно статистике наиболее популярных поисковых запросов Яндекса, во второй половине дня интерес пользователей к работе начинает пропадать, а к 17 часам и угасает окончательно (рисунок 3).

В это время становятся более актуальны игры и прочие развлечения. На практике это означает, что многие сотрудники отвлекаются от работы, а их эффективность понижается.

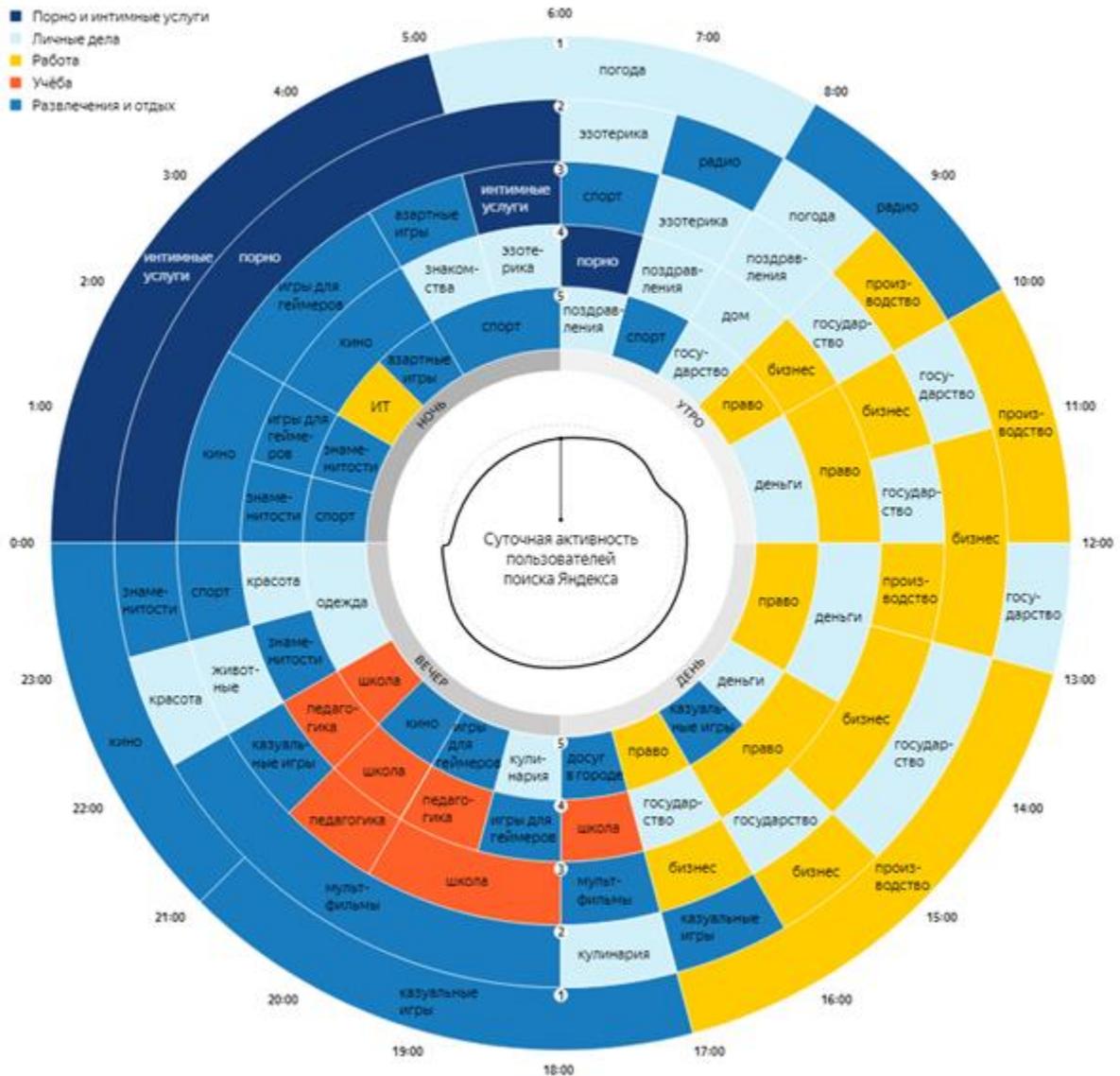


Рис.3. Статистика поисковых запросов Яндекса

Источник: [3]

По данным исследования Зарплата.ру [4], каждый восьмой сотрудник старается делать перерывы от работы как можно чаще. У 33% респондентов на отдых в рабочее время уходит около 30 минут. Еще 30% ответили, что отдыхают около часа рабочего времени. 10% опрошенных признались, что занимаются своими делами около двух часов.

Кроме того, офисные сотрудники в большинстве согласились, что считают гаджеты наилучшим способом отдохнуть в течение рабочего дня.

75% ответили, что проводят перерывы в социальных сетях или на прочих развлекательных веб-сайтах.

Однако, если секретарь вместо того, чтобы отвечать на звонки, переписывается с подругами в социальных сетях или менеджер играет в онлайн игры вместо работы с заказчиками, то такой работник приносит убытки компании. Несложно рассчитать, что, если средняя зарплата в компании составляет около 40 000 рублей, то часовая ставка равна приблизительно 230 рублей. На одного человека, который потратил два часа рабочего времени на развлечения в сети, расходы составят порядка 10 000 в месяц, то есть четверть заработной платы. В крупной компании, где работают сотни человек, расходы будут исчисляться миллионами.

С учетом постоянно растущего уровня конкуренции, руководство любой организации стремится оптимизировать каждую минуту рабочего процесса, чтобы достичь успеха.

Контроль персонала необходим как руководителю, так и сотрудникам. Руководитель получает высокие результаты, лояльность персонала и надежную команду. Сотрудники - рост компании, улучшение ее имиджа влечет за собой стабильную зарплату, мотивацию, прозрачные требования и правила.

Чего позволяет добиться управленческий контроль:

- Сотрудники понимают, что от них требуется благодаря прозрачным KPI
- Контроль персонала способствует высокому качеству и четким срокам работы, что повышает конкурентоспособной компании
- Отлаженное управление обеспечивает стабильность, развитие и способность преодоления кризисных ситуаций
- Ошибки и проблемы могут быть выявлены на ранней стадии, пока они не превратились в критичные, что снижает число убытков

- Четкие правила повышают лояльность персонала. «Строгий, но справедливый» руководитель, как правило, пользуется уважением подчиненных

- Правильно выстроенный контроль в первую очередь направлен на учет успехов, а не провалов, что мотивирует и способствует профессиональному росту персонала

Кроме того, правильно выстроенный контроль сотрудников:

- Выявляет слабую мотивацию и ее причины;
- Способствует нахождению слабых мест в планировании рабочих процессов;

- Позволяет определить, кто из сотрудников не уделяет должного внимания и времени работе и из-за этого срывает сроки;

- Способствует снижению утечки ресурсов компании;

- Помогает «вычислить» сотрудников, работающих с конкурентами.

Основные функции контроля:

- Мотивация: материальное и нематериальное поощрение;

- Принуждение: материальные и нематериальные санкции;

- Дисциплина: контроль способствует повышению ответственности и собранности.

Что касается правовой стороны, всем известно, что законодательство гарантирует неприкосновенность частной жизни. Однако, технически при заключении трудового договора работодатель арендует личное время своих сотрудников. Поэтому оно теряет статус личного и становится рабочим. Следовательно, соответствующая конституционная статья на него уже не распространяется. Кроме того, корпоративный трафик принадлежит компании и все, что через него проходит, тоже. Это относится и к аппаратным средствам, установленным программам и т.д. Поэтому

контроль деятельности сотрудников на рабочем компьютере находится в рамках закона.

## 1.2. ПРЕИМУЩЕСТВА ВНЕДРЕНИЯ ПО ДЛЯ МОНИТОРИНГА СОТРУДНИКОВ

Первая программа для слежения за персоналом появилось в 1888 году. Прогресс не стоит на месте, и по статистике, в настоящее время каждая третья компания мониторит деятельность своих сотрудников за рабочими компьютерами – от просмотра электронной почты и установки тайм-трекеров до тотального контроля трафика.

Программы учета автоматизируют контроль за персоналом. Они учитывают время работы и простоя, установленные приложения и открытые файлы, историю веб-запросов и даже считывают нажатые клавиши, и все это незаметно для сотрудника. Благодаря подобным сервисам руководителям легче анализировать загруженность подчиненных и планировать работу.

Преимущества систем учета рабочего времени и контроля персонала (СУВР):

1) Снижение расходов, связанных с простоем. Подобные ПО позволяют узнать время, когда сотрудник действительно работает, и когда нет, что может быть учтено при начислении зарплаты.

2) Получение большего количества информации о работниках, обнаружение их ошибок на ранних этапах и выявление наиболее перспективных и трудолюбивых.

3) Более эффективное управление командами способствует их слаженной работе

4) Лучшее понимание сильных и слабых сторон каждого способствует более эффективному делегированию полномочий и распределению задач

5) Автоматизация административной работы по учету рабочего времени, устранение ошибок, связанные с человеческим фактором.

### 1.3. ОБЗОР СУЩЕСТВУЮЩИХ СИСТЕМ

В ходе работы были рассмотрены наиболее популярные системы контроля персонала и кейлоггеры. Было выделено пять лучших программ и проведен их сравнительный анализ. Результаты представлены в таблице 1.

Таблица 1.

Сравнительный анализ систем для мониторинга сотрудников

Название	Kickidler	Iwantsoft Free Keylogger	Staffcop	CrocoTime	Yaware
Стоимость	\$54/пк в год	\$203	\$40/пк в год	\$40/пк в год	\$58/пк в год
Запись нажатых клавиш	+	+			
Запись веб- истории		+		+	+
Веб-история со временем					+
Мониторинг чатов			+		
Открытые приложения	+	+	+	+	+

Учет рабочего времени	+		+	+	+
Запись видео экрана	+				
Онлайн просмотр экрана	+		+	+	
Интенсивность нажатия клавиш	+				
Единая временная шкала	+				+
Определение продуктивности открытых окон на основании датасетов	+		+	+	+
Невидимый режим		+		+	
Скриншоты экрана		+	+	+	+

Источник: составлено автором

Итогом сравнительного анализа существующих систем для мониторинга работы персонала являются следующие выводы:

1) Для записи посещаемых веб-ресурсов большинство систем используют историю браузера, в которой нет информации о времени, которое человек провел на каждой странице.

Такой подход не является объективным, так как случайно открытые вкладки оказываются наравне с теми, на которых было проведено большое количество времени, что может привести к неверным выводам.

2) При категоризации интернет-ресурсов берутся не конкретные страницы, которые посещал пользователь, а сайты.

Однако, существуют сайты разной направленности. Например, на youtube можно смотреть как развлекательные ролики, так и гайды по программированию.

3) Для определения категорий веб-страниц все рассмотренные системы используют заранее созданные датасеты. В случае, если интересующего ресурса нет в таком датасете, он помечается как неизвестный или просто игнорируется.

Однако, согласно статистике, в мире насчитывается около 1,8 миллиарда действующих сайтов, и их число растет с каждой секундой [5]. Поэтому, данный способ - не самый эффективный и требует постоянного обновления датасетов. В противном случае, подобный анализ и категоризация теряют смысл.

4) Единый подход для всех сотрудников при определении “продуктивности” ресурса.

Хотя для разных групп специалистов в разных организациях “продуктивные” и “непродуктивные” ресурсы могут отличаться. Например, основной инструмент маркетологов - социальные сети, которые, как правило, помечаются как непродуктивный ресурс.

## ГЛАВА 2. РАЗРАБОТКА ТРЕБОВАНИЙ И ПРОЕКТИРОВАНИЕ

### 2.1 ВЫЯВЛЕНИЕ ТРЕБОВАНИЙ

#### 2.1.1. Цели разработки

Целью разработки является создание системы STAMOS (Staff Monitoring System) для мониторинга и анализа действий пользователей на персональных компьютерах, которая будет лишена выявленных ранее недостатков, благодаря:

- классификации веб-ресурсов с помощью машинного обучения;
- классификации конкретных страниц, а не сайтов;
- определению продолжительности пребывания на веб-странице;
- классификации веб-страниц по тематикам, которые затем будут определяться как “продуктивные” или “непродуктивные” для конкретных групп сотрудников.

#### 2.1.2. Варианты использования системы

На этапе разработки и анализа требований к системе STAMOS (Staff Monitoring System) были определены ее основные функции, отображенные на диаграмме вариантов использования (рисунок 4).

Система состоит из двух подсистем – десктопного приложения для сотрудника и веб-приложения для менеджера.

С системой взаимодействуют два актера - сотрудник, за действиями которого осуществляется контроль, и менеджер, контролирующий его.

Менеджер создает в приложении рабочую группу и добавляет внутри нее сотрудников. По каждому сотруднику он может посмотреть статистику его активности в течение рабочего времени. Также он может выбрать, какие категории веб-страниц считать продуктивными для конкретной рабочей группы.

Для сотрудника единственным прецедентом является авторизация в приложении для начала мониторинга и выход из учетной записи для его остановки.



Рис.4. Варианты использования системы

Источник: составлено автором

### 2.1.3. Потоки данных системы

Для определения того, как данные будут перемещаться между процессами, хранилищами и внешним миром была разработана диаграмма потоков данных (Data Flow Diagram).

Контекстная диаграмма DFD (рисунок 5) – модель окружения – представляет всю систему как единый процесс, изображенный в виде одного блока. Внешней сущностью являются менеджер, сотрудник, операционная система, браузер и платформа машинного обучения Azure machine learning studio.



Рис.5. Контекстная диаграмма DFD

Источник: составлено автором

Диаграмма декомпозиции контекстной диаграммы потоков данных – модель поведения, представленная на рисунке 6, конкретизирует контекстную диаграмму, выделяя важнейшие процессы системы. На диаграмме декомпозиции отображаются все внешние сущности и потоки данных контекстной диаграммы. Кроме того, здесь также показываются хранилища данных для моделирования данных, которые должны сохраняться в промежутках между процессами.



Рис.6. Диаграмма декомпозиции DFD

Источник: составлено автором

Как видно из диаграммы, основными процессами внутри системы являются:

- Добавление менеджером нового сотрудника
- Авторизация сотрудника
- Обработка данных о веб-страницах, полученных от браузера
- Извлечение текста со страниц
- Определение категорий веб-страниц
- Формирование истории открытых приложений
- Формирование единой статистики сотрудника

## 2.2. ПРОЕКТИРОВАНИЕ АРХИТЕКТУРЫ

Диаграмма развертывания системы представлена на рисунке 7.

Она состоит из персонального компьютера, на котором установлено приложение STAMOS, сервера с брокером очередей RabbitMQ, веб-сервера, содержащего веб-приложение и обслуживающие сервисы,

персонального компьютера с браузером, из которого открывается веб-приложение и сервера базы данных с СУБД MySQL Server.

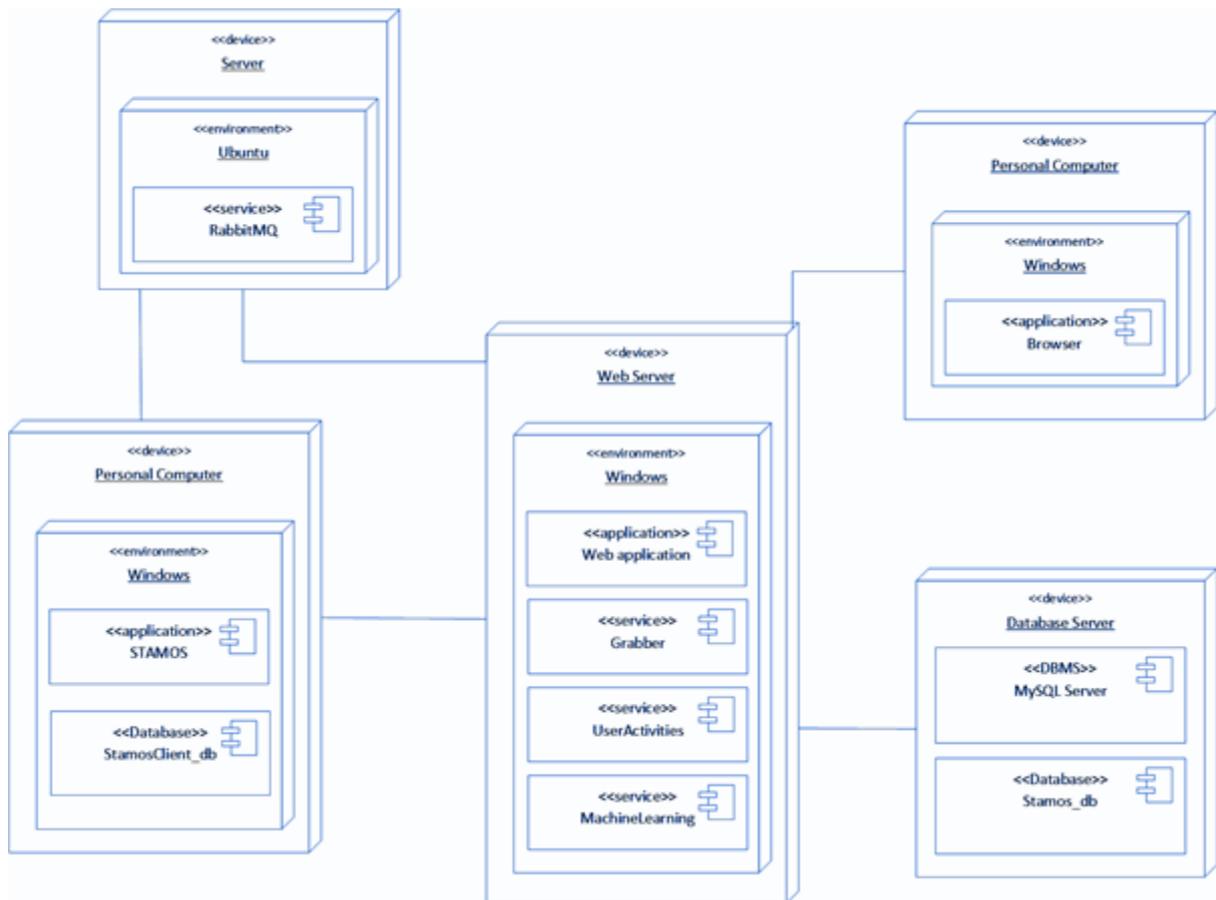


Рис.7. Диаграмма развертывания

Источник: составлено автором

### 2.3. ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ

Были спроектированы две базы данных - сервера и клиента.

Клиентская база данных состоит из одной таблицы, содержащей данные о пользовательских аккаунтах. Для ее реализации было принято решение использовать SQLite ввиду ее легковесности. Физическая модель представлена на рисунке 8.

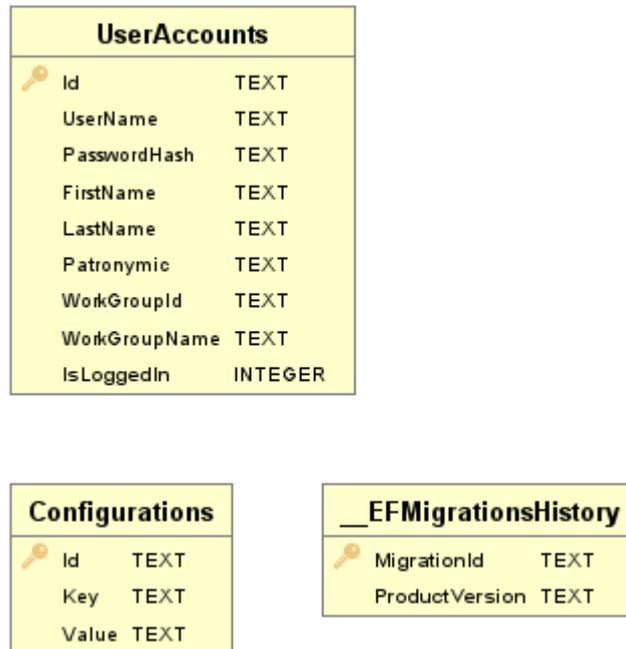
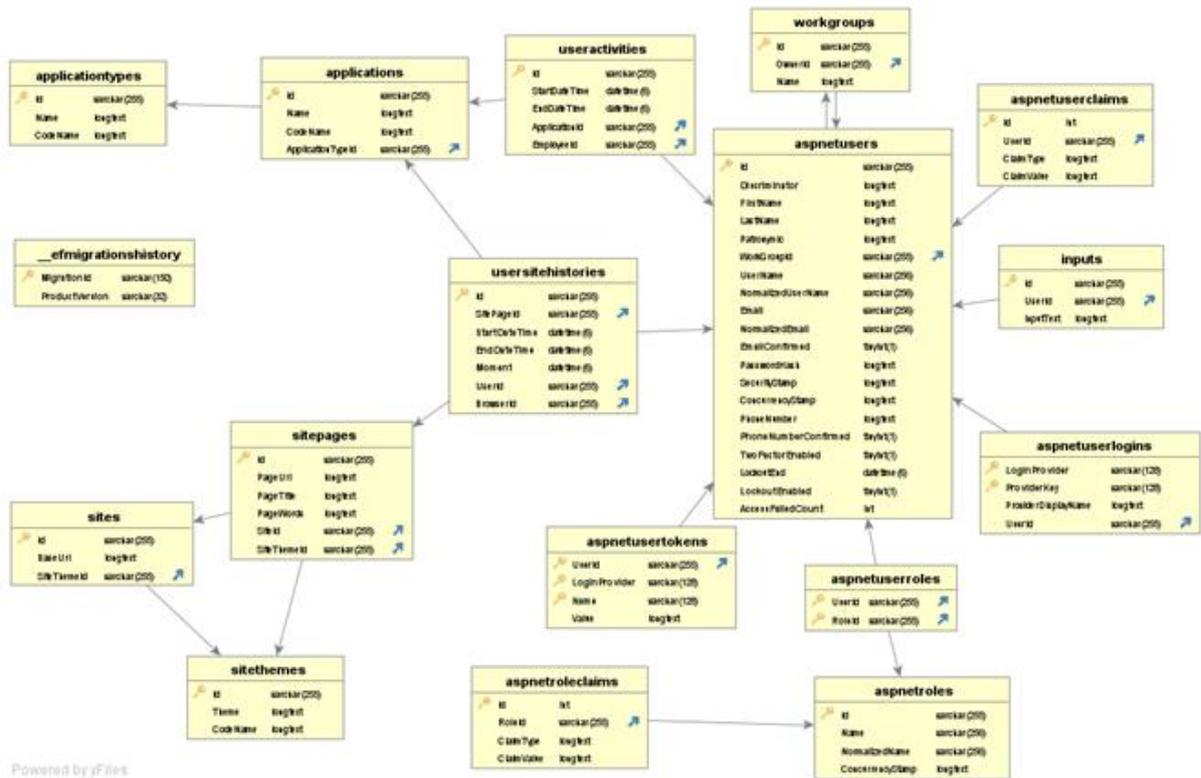


Рис.8. Клиентская база данных

Для серверной базы данных используется СУБД MySQL Server.

Физическая модель представлена на рисунке 9.

Главной сущностью здесь так же является пользователь. Причем она объединяет в себе два типа пользователей - менеджер и сотрудник, что реализовано благодаря механизму наследования. Менеджер является владельцем рабочих групп, в которых состоят сотрудники. У сотрудников есть их история активности в приложениях и на веб-ресурсах.



Powered by yFiles

Рис.9. Серверная база данных

## ГЛАВА 3. КОНСТРУИРОВАНИЕ СИСТЕМЫ

### 3.1. ИНСТРУМЕНТЫ РАЗРАБОТКИ

Основным языком разработки был выбран C#, в первую очередь ввиду удобной разработки под Windows.

Для разработки веб-приложения был выбран фреймворк ASP.NET 5.

Для десктопного приложения - технология Windows Forms. Причинами такого решения являются, во-первых, возможность минимизации приложения в системный трей при попытке закрытия, что крайне проблематично реализовать, например, при использовании WPF. Во-вторых, разрабатываемое приложение не требует больше никакой особенной, гибкой настройки и не предполагает сложных дизайнерских решений, ввиду чего нет необходимости в использовании более сложных технологий.

Для машинного обучения система взаимодействует с веб-порталом Microsoft Azure Machine Learning Studio. Данный сервис отличается широким функционалом и удобным интерфейсом. Кроме того, важным фактором является возможность взаимодействия через API.

Кроме того, система использует брокер очередей RabbitMQ, снижающий нагрузку на веб-приложение и обеспечивающий высокую степень отказоустойчивости системы, необходимую при взаимодействии с внешними системами - анализируемыми веб-сайтами и порталом машинного обучения ввиду вероятности их недоступности.

### 3.2. РЕШЕНИЯ ПО РЕАЛИЗАЦИИ

Получение данных об открытых приложениях происходит следующим образом: берется список процессов, открытых на ПК, после чего записываются названия главных окон у тех, у которых оно есть (именно они нас интересуют, приложения без главных окон - сервисы).

Подобная процедура забора и отправки данных об открытых приложениях происходит каждые 30 секунд. Если приложения, поступившего в текущий забор, не было в предыдущем или было проставлено время закрытия - оно записывается в историю с текущим временем как временем открытия. Если приложения, которое было открыто в предыдущий раз, нет в поступившем списке - для него записывается текущее время как время закрытия.

История открытых вкладок добывается из внутренней базы данных браузера. На данный момент рассматривается Google Chrome и его SQLite база данных (рисунок 10), которая копируется в директорию приложения, где для нее создается контекст и модели. Информация о вкладках и времени формируется из таблиц `visits` и `urls` и отправляется на сервер, где обрабатывается. Новые данные, если они есть, добавляются в базу. Запрос отправляется раз в 10 секунд.

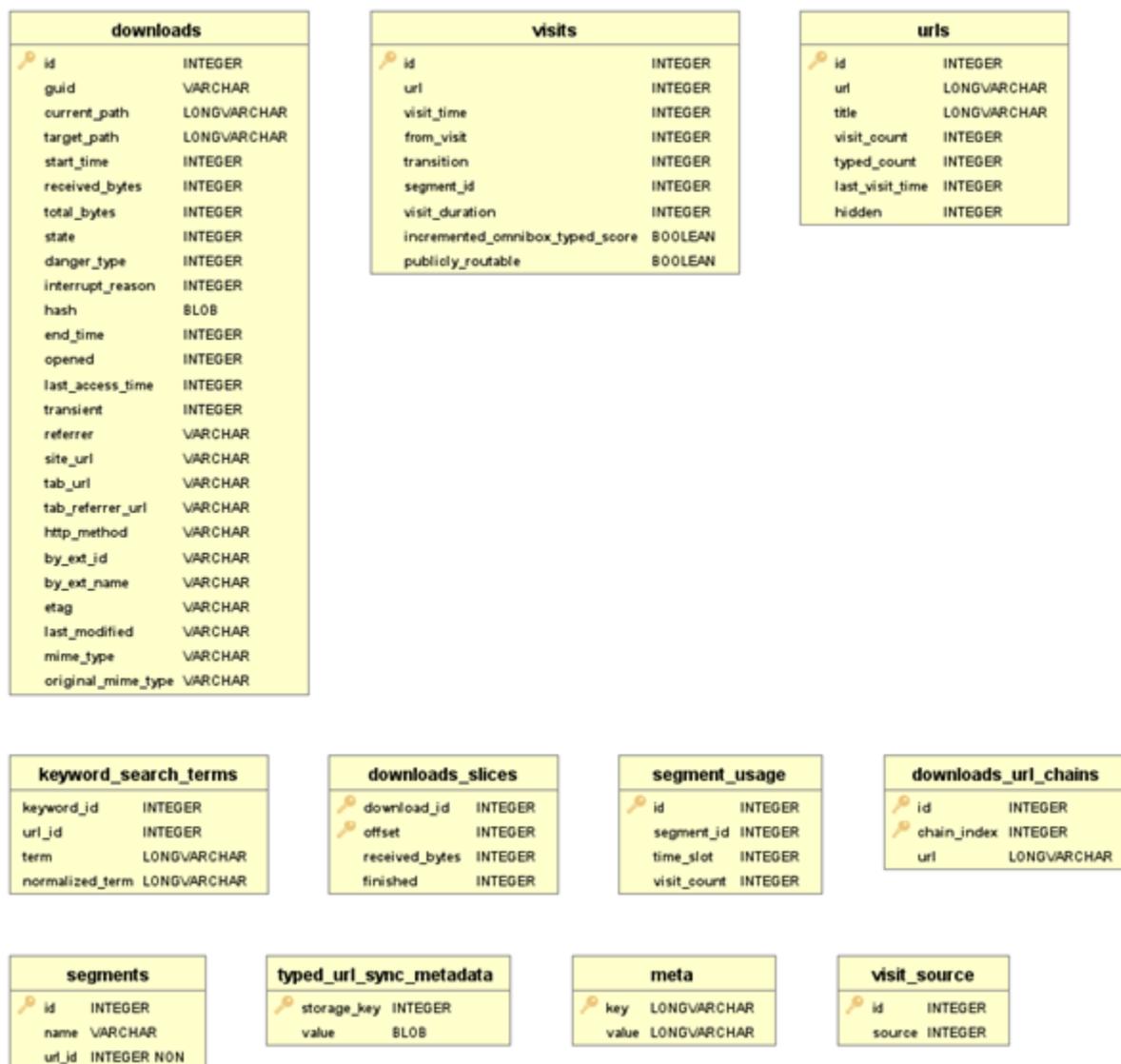


Рис.10. База данных Google Chrome

Аутентификация пользователя в десктопном приложении происходит через обращение к API веб-сервера.

### 3.3. МАШИННОЕ ОБУЧЕНИЕ ДЛЯ ОПРЕДЕЛЕНИЯ КАТЕГОРИИ ВЕБ-СТРАНИЦЫ

Для реализации машинного обучения в первую очередь был сформирован датасет из веб-страниц.

Для классификации страниц было выделено 7 категорий:

- Развлечения;

- Новости;
- Образование;
- Сайты для взрослых;
- Сайты для поиска работы;
- Интернет-магазины;
- Социальные сети.

Далее были вручную отобраны страницы и определены их категории.

После отбора веб-страниц, сервис Grabber забирает их содержимое.

Контент веб-страницы состоит из следующих блоков:

- Название;
- Заголовки (<h1> <h2> и т.д.);
- Meta keywords;
- Meta description;
- Текстовый контент (<p>, <span>, <label> и т.д.);
- Контент, описывающий особенности функционирования или отображения (<script>, <style>, <audio>, <img> и т.д.).

Последний блок вырезается, так как не содержит значащего контента.

В итоге датасет составляется следующим образом: отдельно записывается название страницы, все заголовки, которые присутствуют на странице, в отдельные поля записываются meta keywords и meta description, и отдельно весь остальной текст. Пример нескольких записей в базе данных такого вида представлен на рисунке 11.

Url	Title	MetaDescription	MetaKeywords	Headings	Text	Themeld
https://www.postgresqltutorial.com/p...	PostgreSQL CREATE TABLE	this tutorial shows you how...		postgresql create table postgresql create ...	skip to content home getting started advanced views indexes tr...	ed6299fb-c606-4c1d-adca-0b7b2dcbce363
https://www.ea.com/games/library/...	Multiplayer Video Games - ...	explore multiplayer video g...		ea multiplayer games	ea play fifa madden nfl apex legends star wars squadrons the si...	30749406-ae16-491a-bd27-1c630d0d713a
https://www.theladders.com/job/tax-...	TAX Accountant - Salem   La...	this company is hiring a tax...	tax accountant highest paying jobs	tax accountant in salem ma job descripti...	jobs locations companies titles remote jobs nyc jobs milwaukee...	1bc8e451-c7b1-4074-9ee5-039c4ba2074b
https://www.tiktok.com/@khabylame...	Khabane lame (@khabylame)	khabane lame khabylame ...	khabane lame khabylame tiktok ...	for you following khabylame khabane la...	tiktok upload log in keyboard shortcuts go to previous video go...	1c9022e0-7cfe-412d-ad4f-a3e11137e054
https://www.theladders.com/jobs/ma...	Highest Paying Marketing ...	search marketing media de...		filter your results marketing media desig...	jobs locations companies titles remote jobs nyc jobs milwaukee...	1bc8e451-c7b1-4074-9ee5-039c4ba2074b
https://www.monsterindia.com/job/...	Walk-In Interviews for Degre...	job description for walk in i...		walk in interviews for degree freshers bpo...	jobseeker login job search jobs by location jobs in bangalore jo...	1bc8e451-c7b1-4074-9ee5-039c4ba2074b

Рис.11. Пример нескольких записей для обучения

С помощью сервиса Azure было обучено 5 классификаторов для каждого типа контента:

- логистическая регрессия;
- нейронная сеть;

- лес решений;
- джунгли решений;
- и один против всех.

Схема обучения представлена на рисунке 12.

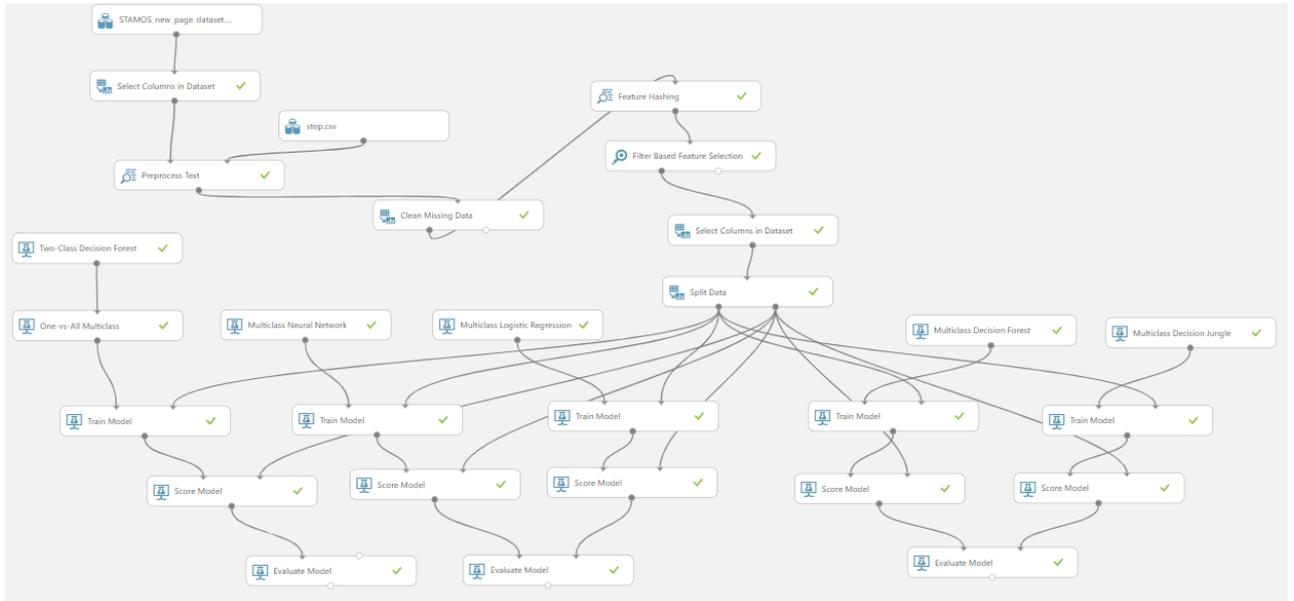


Рис.12. Схема сравнения классификаторов

Результаты обучения классификаторов представлены на рисунках 13-27.

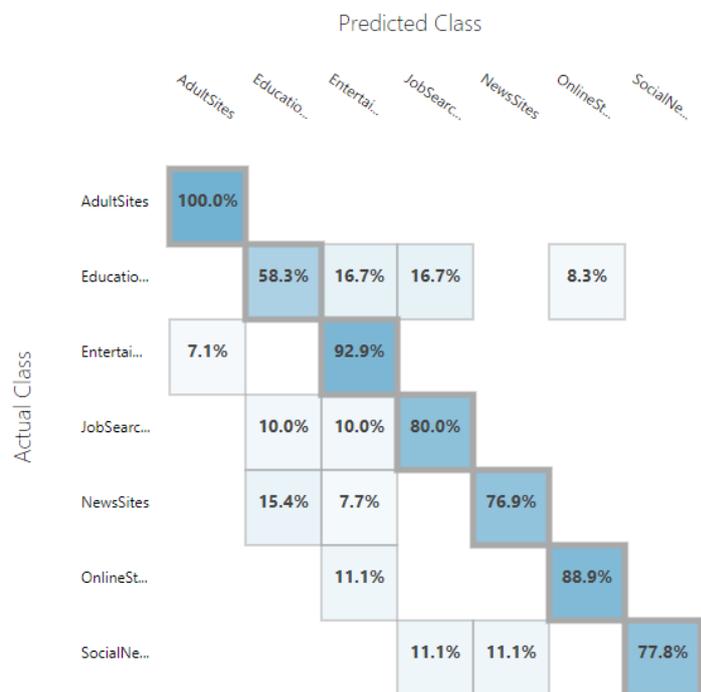


Рис.13. Один-против-всех для текста



Рис.14. Нейронная сеть (слева) и логистическая регрессия (справа) для текста



Рис.15. Лес решений (слева) и джунгли решений (справа) для текста

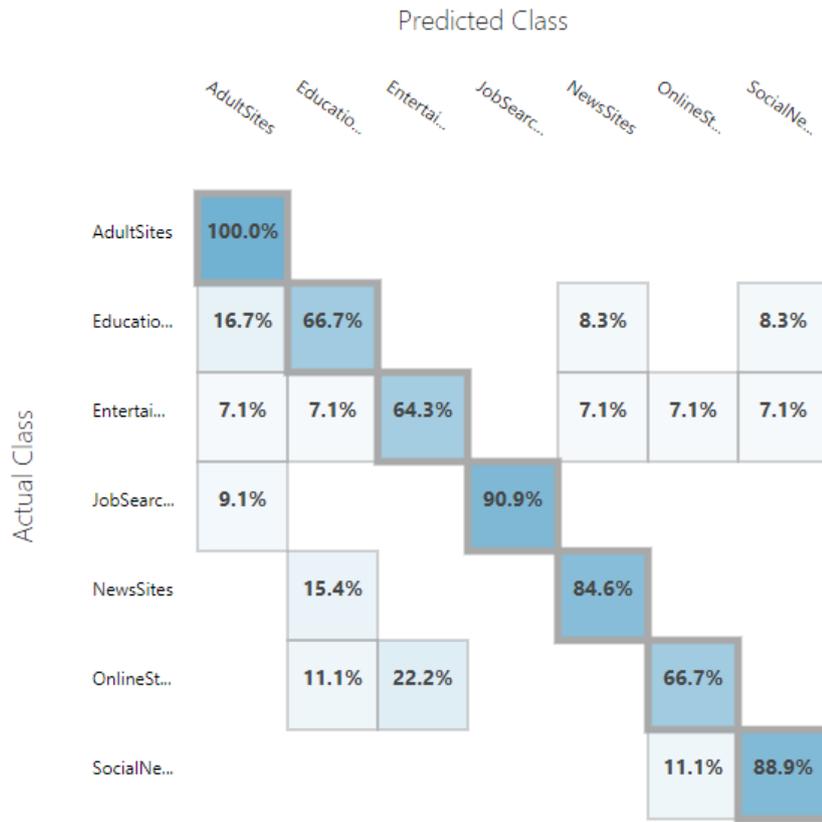


Рис.16. Один-против-всех для названий



Рис.17. Нейронная сеть (слева) и логистическая регрессия (справа) для названий



Рис.18. Лес решений (слева) и джунгли решений (справа) для названий

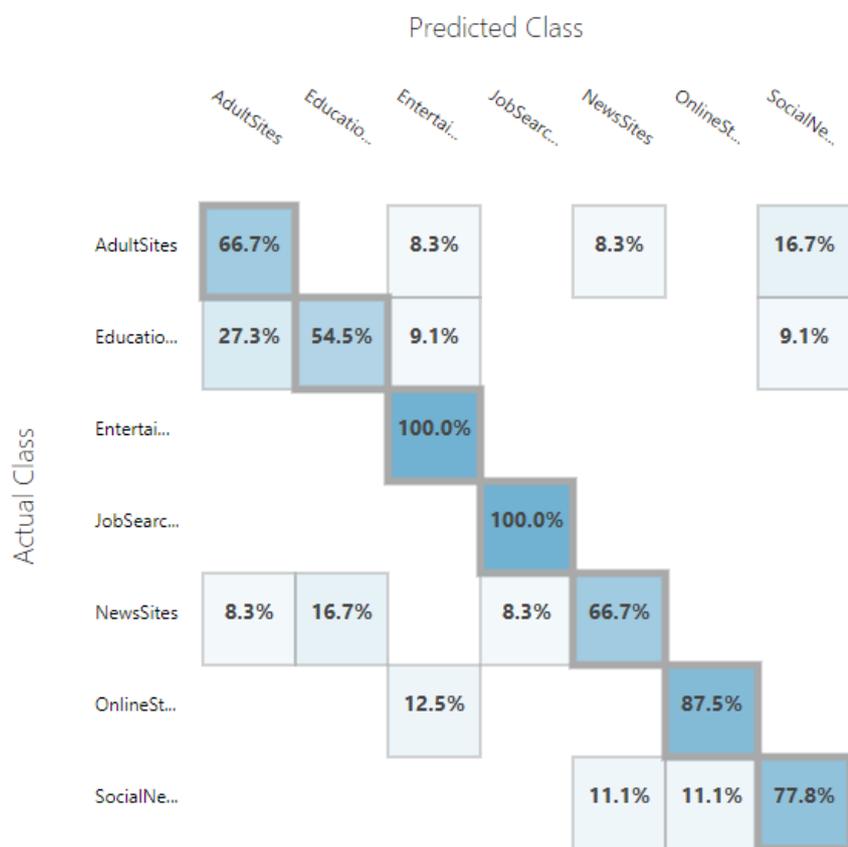


Рис.19. Один-против-всех для заголовков



Рис.20. Нейронная сеть (слева) и логистическая регрессия (справа) для ЗАГОЛОВКОВ



Рис.21. Лес решений (слева) и джунгли решений (справа) для заголовков

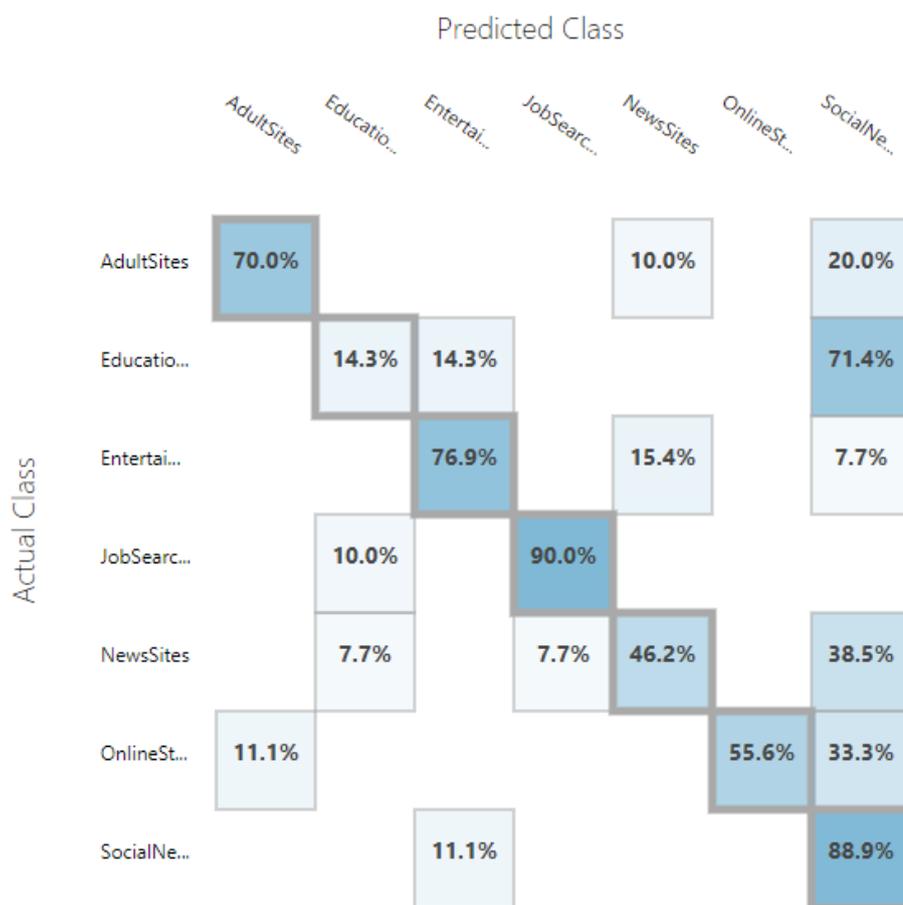


Рис.22. Один-против-всех для meta description



Рис.23. Нейронная сеть (слева) и логистическая регрессия (справа) для meta description



Рис.24. Лес решений (слева) и джунгли решений (справа) для meta description

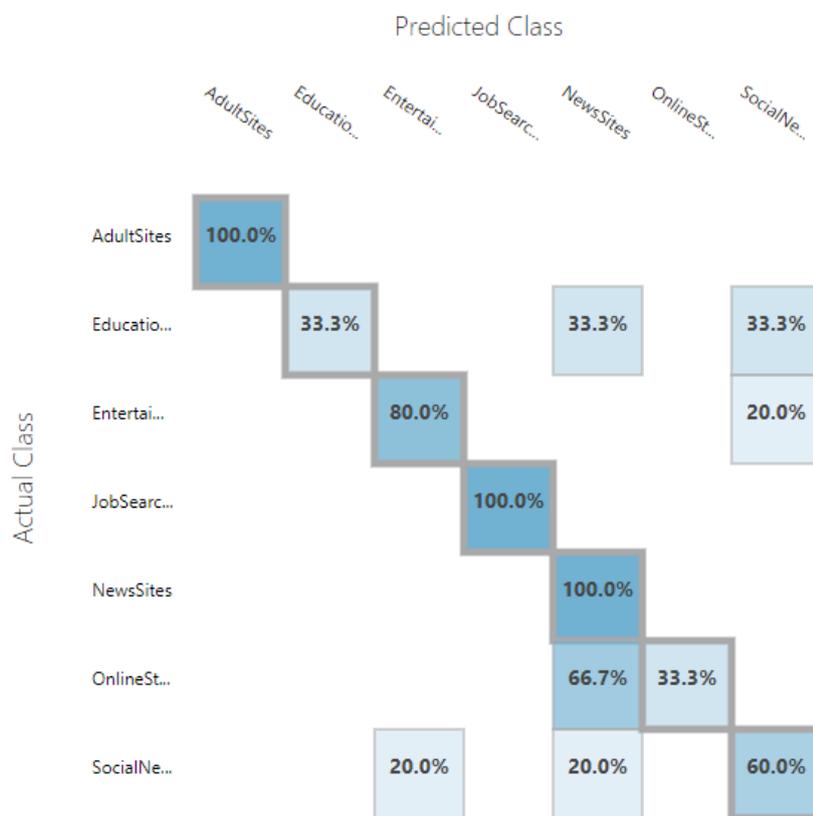


Рис.25. Один-против-всех для meta keywords



Рис.26. Нейронная сеть (слева) и логистическая регрессия (справа) для meta keywords



Рис.27. Лес решений (слева) и джунгли решений (справа) для meta keywords

Для лучшей наглядности результатов была составлена сводная таблица (таблица 2).

Таблица 2.

## Сравнение точности классификаторов

	One-vs-All	Neural network	Logistic regression	Decision forest	Decision jungle
текст	0.822785	0.772152	0.632911	0.734177	0.759494
название	0.8	0.775	0.7125	0.725	0.5875
заголовки	0.783784	0.837838	0.716216	0.689189	0.621622
meta description	0.647887	0.661972	0.549296	0.605634	0.605634
meta keywords	0.758621	0.724138	0.482759	0.724138	0.689655

Как видно из таблицы, классификатор один-против-всех показал наилучшие результаты для текста, названий и meta keywords. А для заголовков и meta description лучшей оказалась нейронная сеть.

Классификатор один-против-всех реализует мультиклассовую классификацию путем трансформации ее в бинарную классификацию. Если дана классификационная проблема с  $N$  возможными решениями, один-против-всех решение состоит из  $N$  отдельных бинарных классификаторов - по одному бинарному классификатору для каждого возможного выхода. В течение тренировки модель проходит через последовательность бинарных классификаторов, тренируя каждый отвечать на отдельный классификационный вопрос.

Для создания веб-сервиса было обучено пять лучших классификаторов для каждого направления - текста, названия страницы, заголовков и двух видов мета-тегов. Результаты по каждой категории представлены на рисунках 28-32.

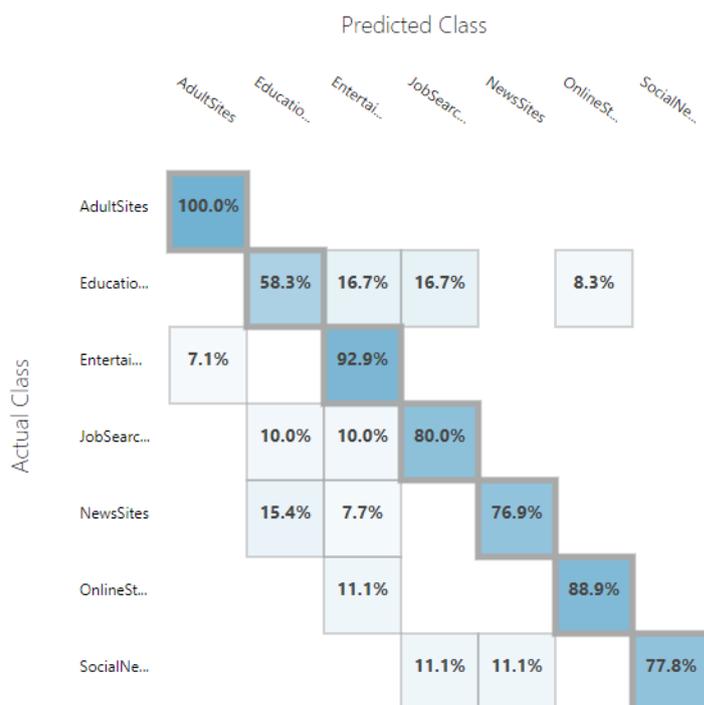


Рис.28. Точность классификатора по основному текстовому контенту

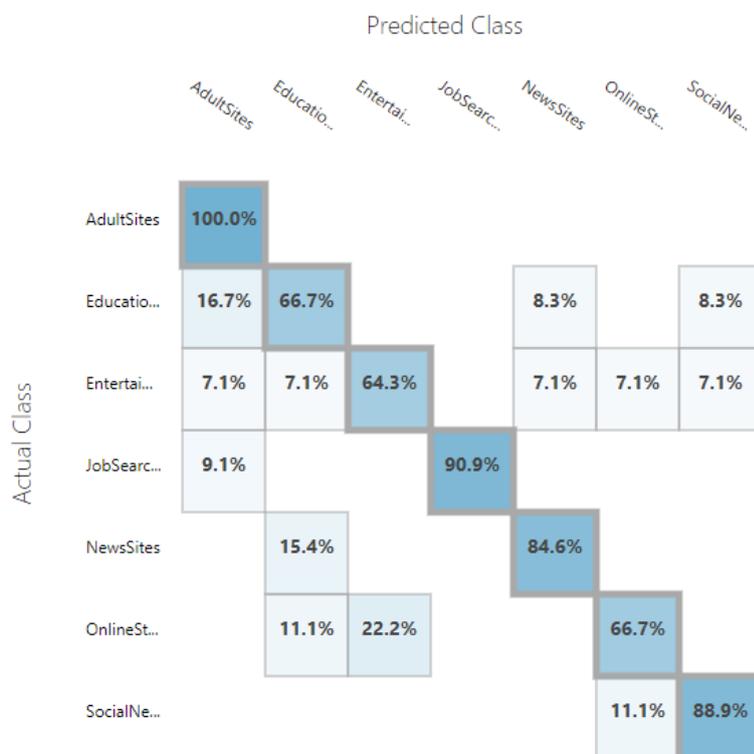


Рис.29. Точность классификатора по названию страницы



Рис.30. Точность классификатора по заголовкам

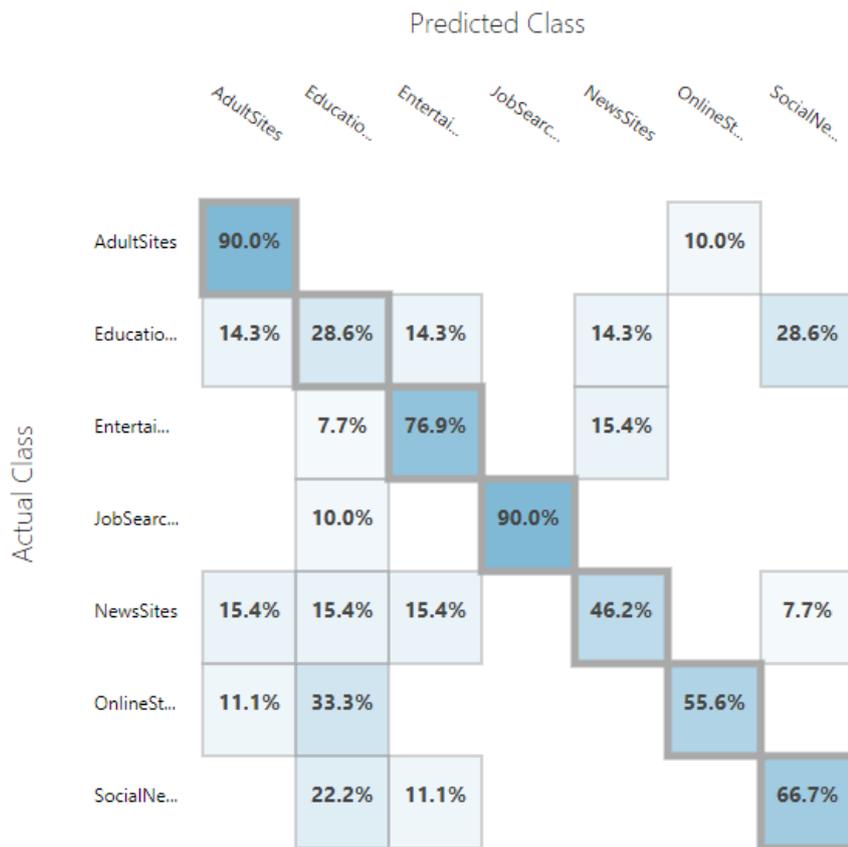


Рис.31. Точность классификатора по meta description

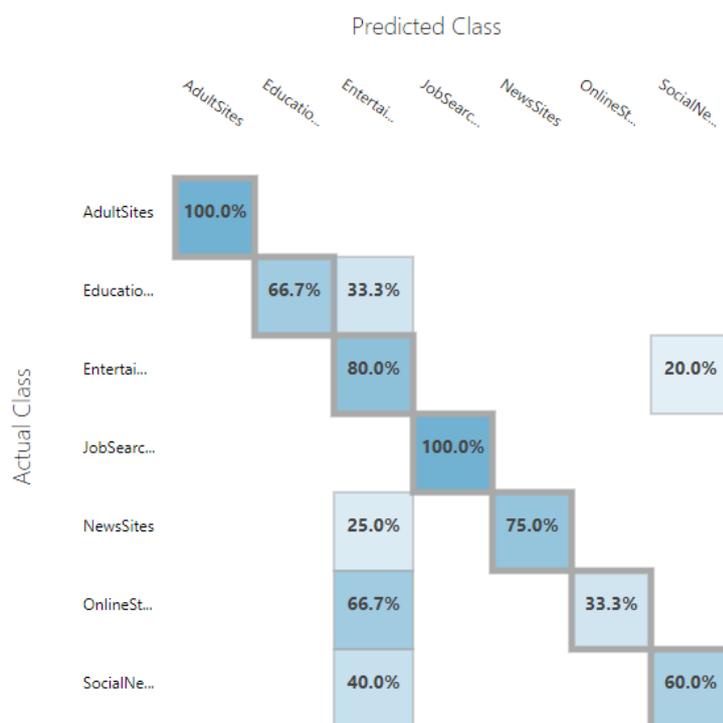


Рис.32. Точность классификатора по meta keywords

Для лучшей наглядности все полученные результаты были объединены в таблице 3.

Таблица 3.  
Результаты обучения

	Категории веб-страниц							
	Для взрослых	Образовательные	Развлекательные	Поиск работы	Новости	Магазины	Социальные сети	Средняя точность
Текст	100%	58.3%	92.9%	80%	76.9%	88.9%	77.8%	82.3%
Название	100%	66.7%	64.3%	90.9%	84.6%	66.7%	88.9%	80%
Заголовки	83.3%	81.8%	100%	100%	75%	87.5%	55.6%	83.8%
Meta description	90%	28.6%	76.9%	90%	46.2%	55.6%	66.7%	64.8%
Meta keywords	100%	66.7%	80%	100%	75%	33.3%	60%	75.9%

По средней точности лучшим оказался классификатор заголовков, однако, как видно из таблицы, разные типы контента на веб-странице по-разному работают для разных категорий страниц. У каждого классификатора есть категории, которые они определяют лучше, и хуже. Причем какой-то классификатор хорошо показывает себя для одной категории, а какой-то - для другой.

Поэтому была сформирована гипотеза о том, что если объединить все классификатором определенным образом, то точность можно повысить. Поэтому на следующем шаге показатели из данной таблицы были использованы в качестве коэффициентов доверия при классификации.

Были созданы веб-сервисы по каждому типу контента, их схему представлены на рисунках 33-37.

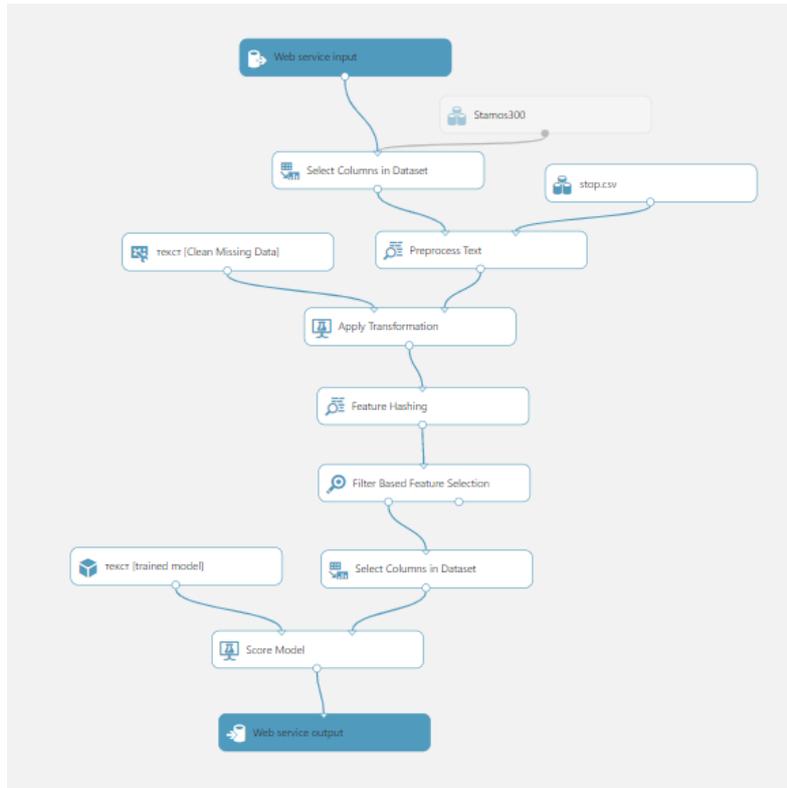


Рис.33. Веб-сервис по тексту

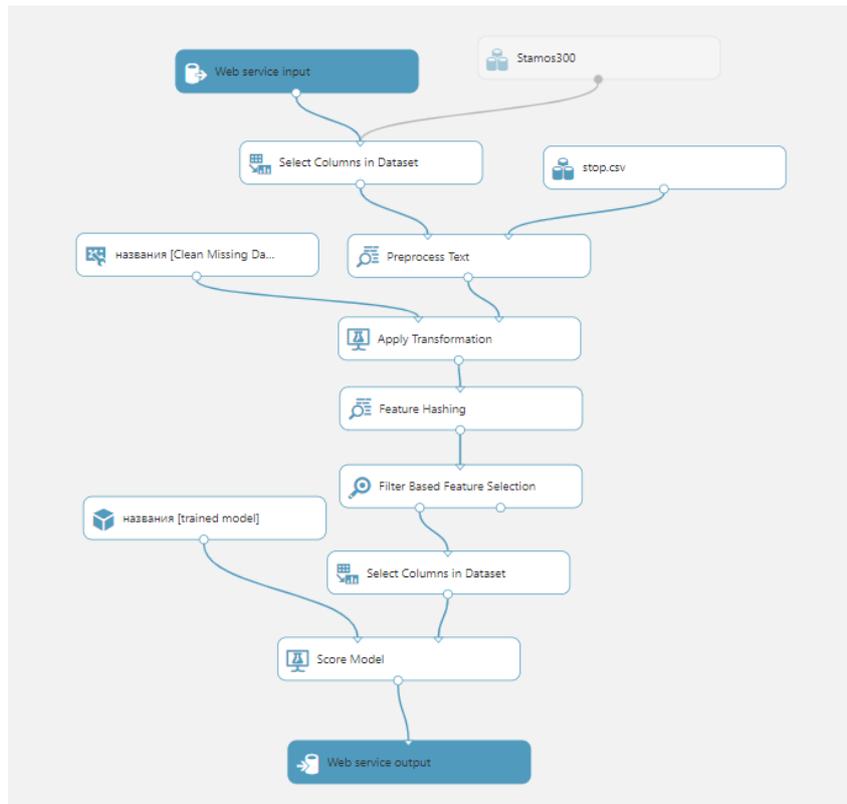


Рис.34. Веб-сервис по названиям

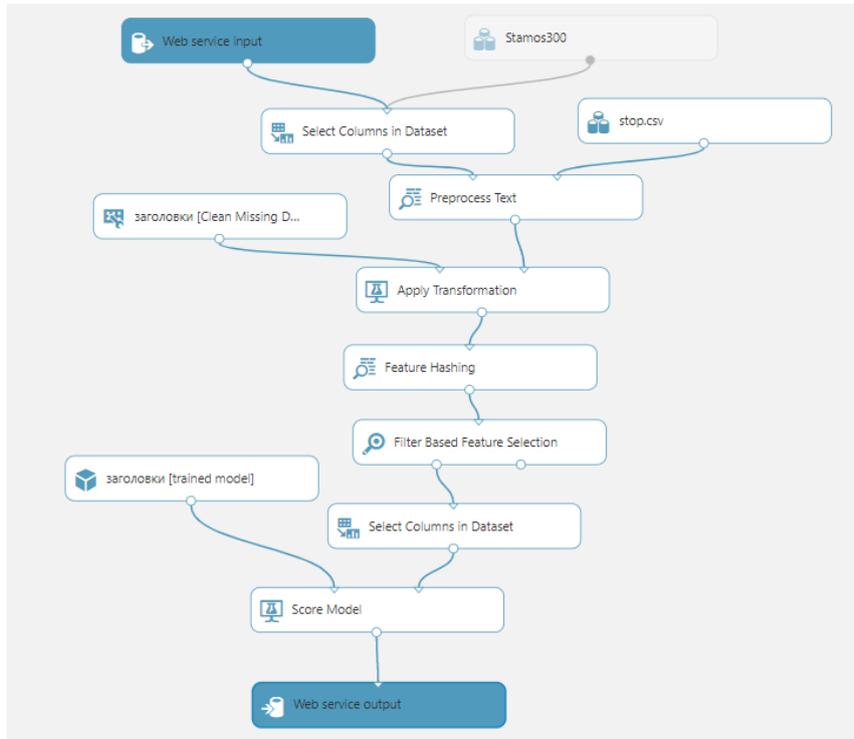


Рис.35. Веб-сервис по заголовкам

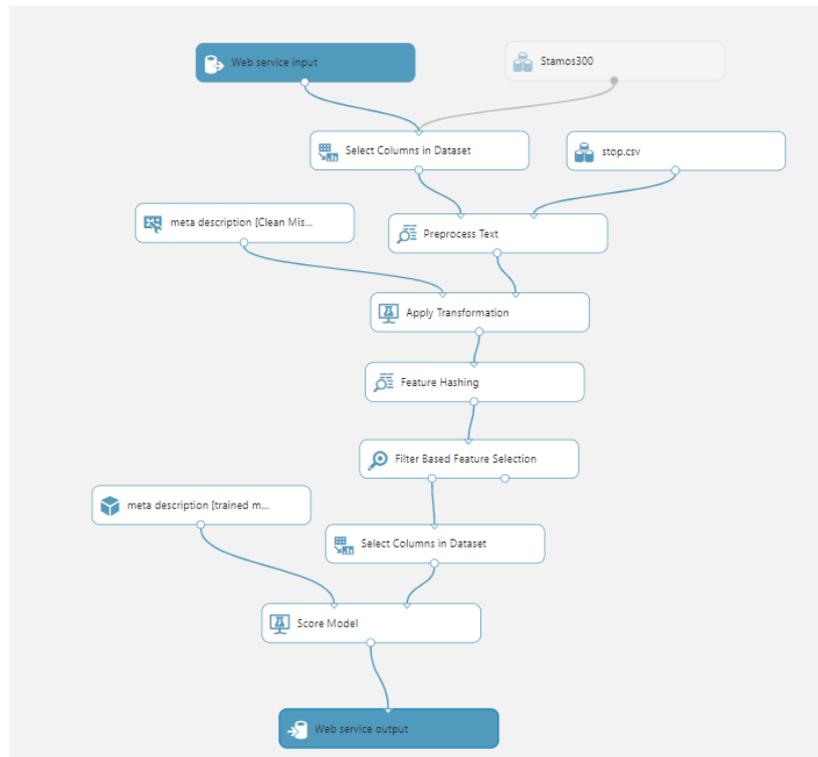


Рис.36. Веб-сервис по meta description

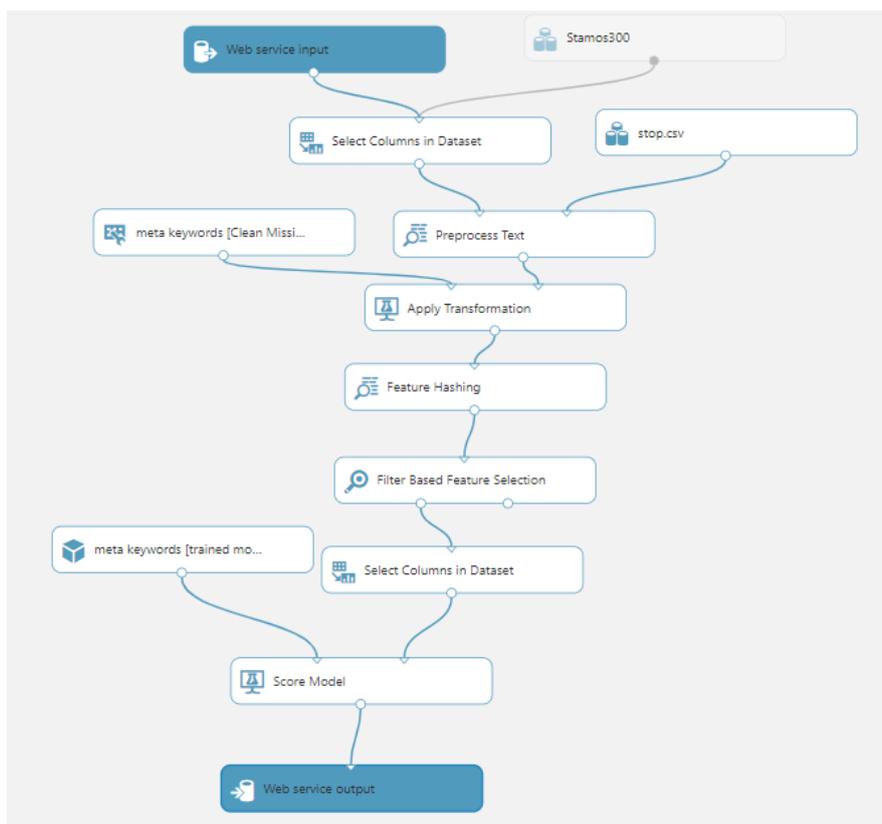


Рис.37. Веб-сервис по meta keywords

После получения вероятностей для каждого класса от всех пяти классификаторов по типу контента, считается оценка степени принадлежности страницы к категории по формуле - вероятности по каждому типу контента умножаются на соответствующие коэффициенты доверия и складываются по каждой категории страниц.

$$R_i = R_{i1} * P_{i1} + R_{i2} * P_{i2} + R_{i3} * P_{i3} + R_{i4} * P_{i4} + R_{i5} * P_{i5}$$

,где  $i$  - номер категории

$R_{ij}$  - коэффициент доверия классификатора  $j$  к категории  $i$

$P_{ij}$  - вероятность категории  $i$ , полученная классификатором  $j$

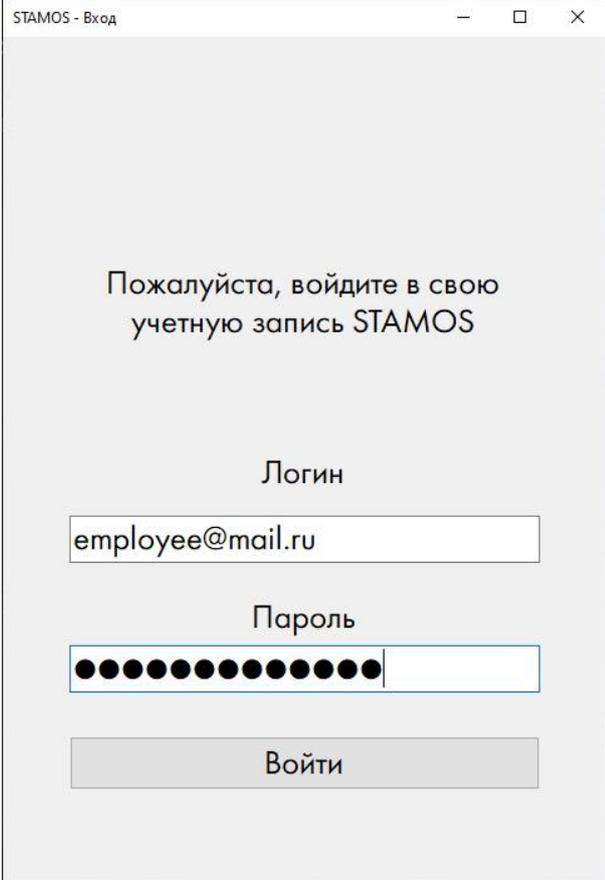
$R_i$  - итоговая оценка для категории  $i$

Итоговая категория веб-страницы определяется как категория с максимальной  $R$ .

Итоговая общая точность классификации, полученная экспериментально - 87.4%.

### 3.4. РАЗРАБОТКА ПОЛЬЗОВАТЕЛЬСКОГО ИНТЕРФЕЙСА

Настольное приложение состоит из двух экранов - входа и приветственного, при закрытии сворачивается в трей (рисунки 38-40).



STAMOS - Вход

Пожалуйста, войдите в свою  
учетную запись STAMOS

Логин

employee@mail.ru

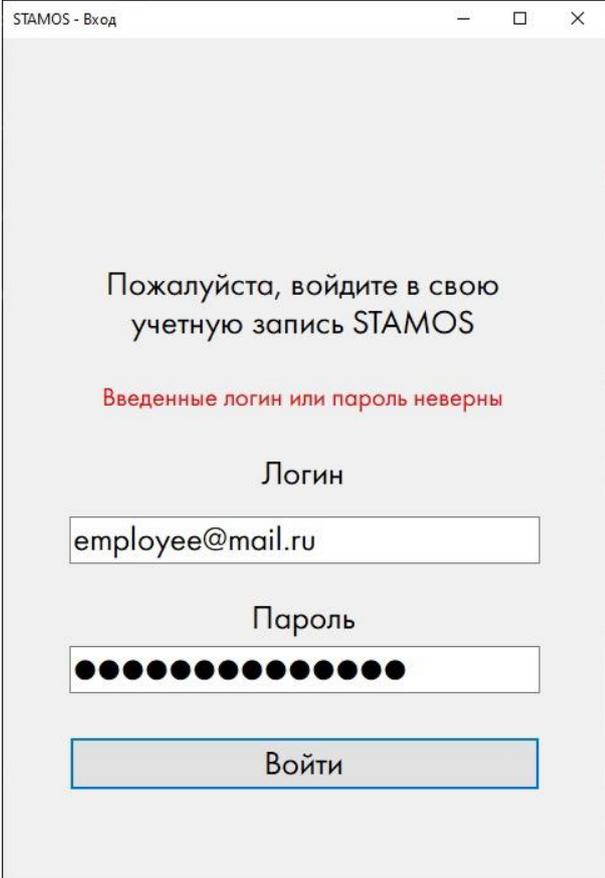
Пароль

●●●●●●●●●●

Войти

Рис.38. Окно авторизации

При неправильных аутентификационных данных выводится сообщение об ошибке.



STAMOS - Вход

Пожалуйста, войдите в свою учетную запись STAMOS

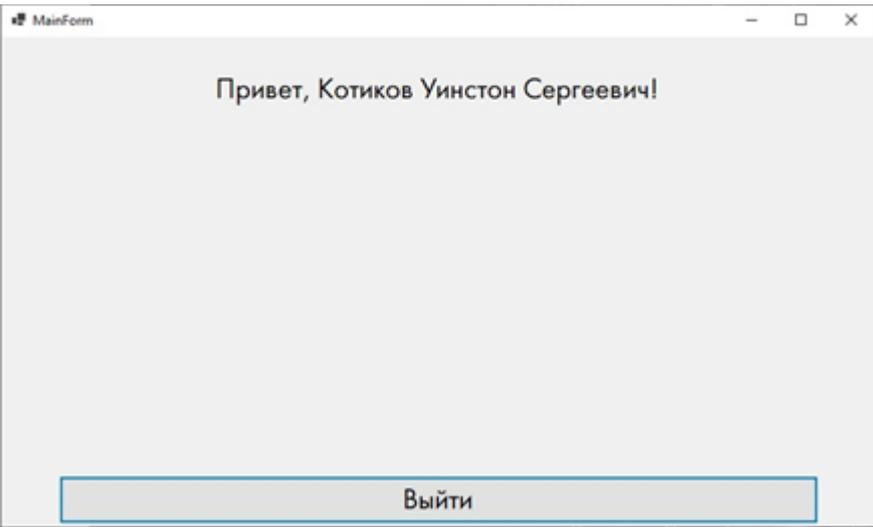
Введенные логин или пароль неверны

Логин

Пароль

Войти

Рис.39. Сообщение об ошибке



MainForm

Привет, Котиков Уинстон Сергеевич!

Выйти

Рис.40. Приветственное окно

У веб-приложения есть страница просмотра рабочих групп и сотрудников, входящих в каждую группу (рисунки 41-42). Есть

возможность добавить новую группу и нового сотрудника, а также отредактировать их данные или удалить (рисунки 43-45).

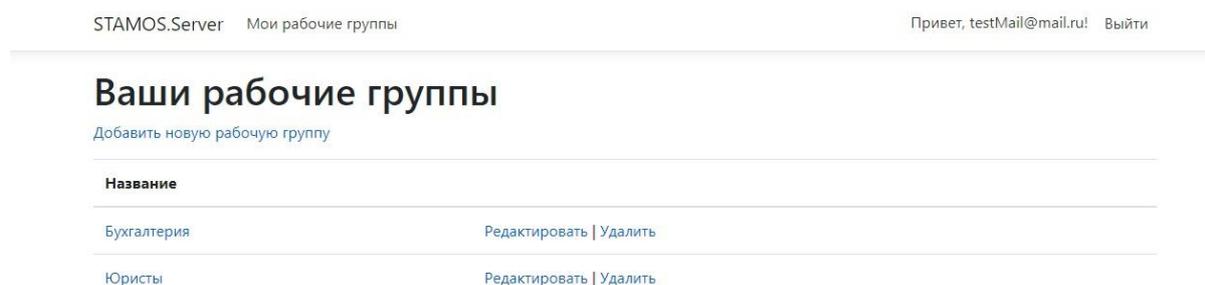


Рис.41. Список рабочих групп

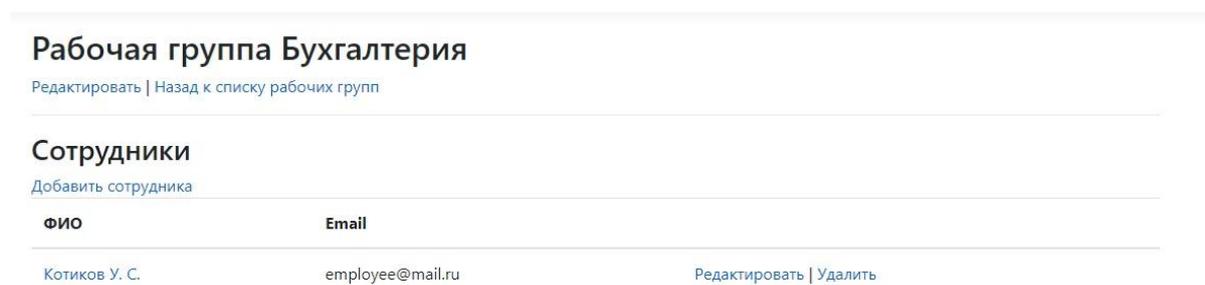


Рис.42. Список сотрудников в группе

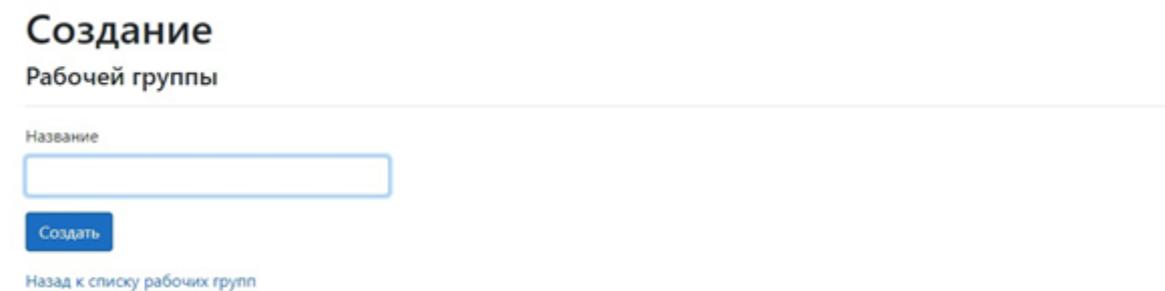


Рис.43. Создание рабочей группы

## Удаление рабочей группы

Вы уверены что хотите удалить данную рабочую группу?

Название: Бухгалтерия

[Удалить](#) | [Назад к списку рабочих групп](#)

Рис.44. Удаление рабочей группы

STAMOS.Server Мои рабочие группы

## Добавить сотрудника

Рабочая группа

Бухгалтерия

Фамилия

Имя

Отчество

Email

Пароль

[Добавить](#)

[Назад к списку сотрудников](#)

Рис.45. Добавление нового сотрудника

Также можно выбрать, какие категории веб-страниц считать продуктивными для конкретной рабочей группы (рисунки 46-47).

## Редактирование рабочей группы

Название

Продуктивные категории

Образовательные сайты × Выберите категории

Сохранить

[Назад к списку рабочих групп](#)

Рис.46. Редактирование рабочей группы

## Редактирование рабочей группы

Название

Продуктивные категории

Образовательные сайты × Выберите категории

- Интернет-магазины
- Новости Press to select
- Поиск работы
- Развлечения
- Сайты для взрослых

Рис.47. Выбор продуктивных категорий

Статистика сотрудников отображается с помощью тайм-лайнов. На рисунке 48 отображена статистика по открытым приложениям.

## Данные по работнику Котиков У. С.

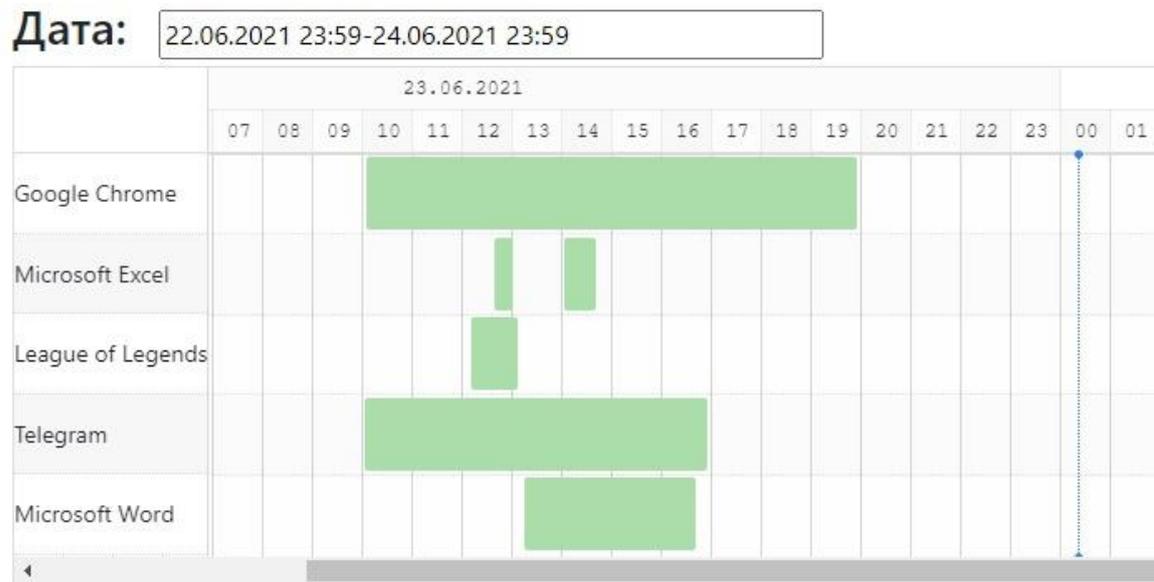


Рис.48. Статистика использованных приложений

При нажатии на строку браузера открывается окно со статистикой по истории посещенных веб-ресурсов (рисунок 49). Выводится основная статистика в виде основных показателей и графиков по категориям и по продуктивности.

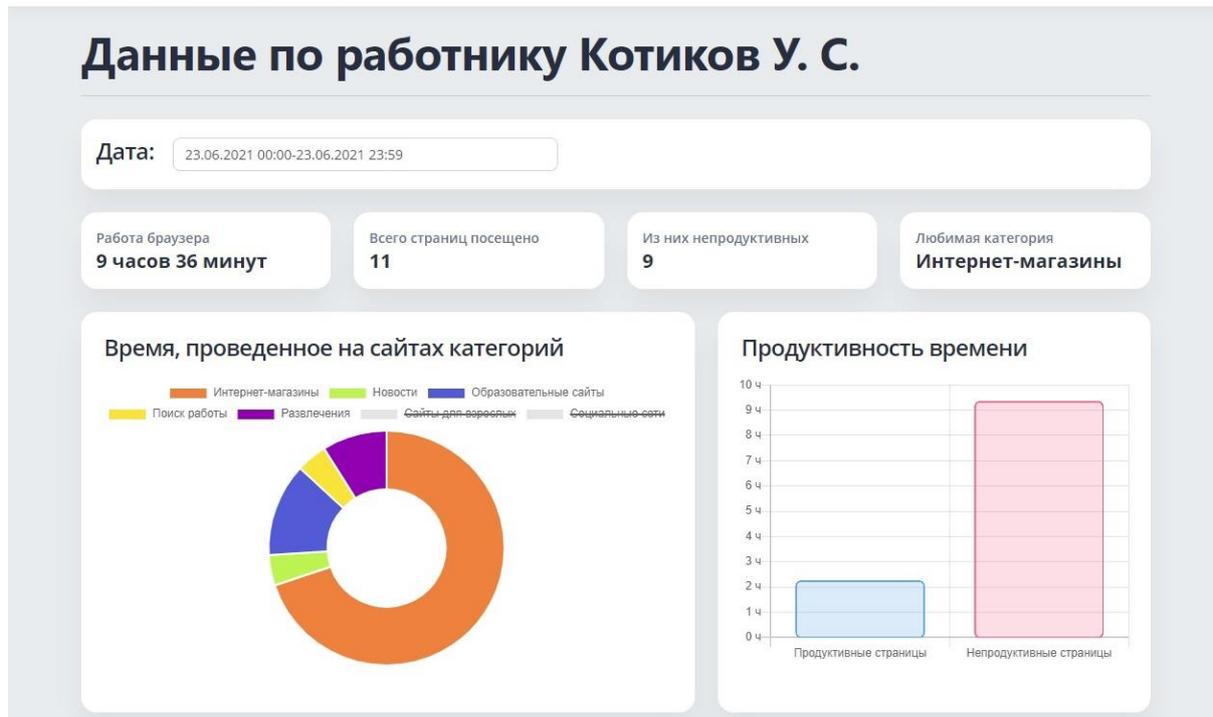


Рис.49. Статистика активности браузера

История посещения веб-ресурсов отображается в виде таймлайна, цветами помечаются разные категории (рисунок 50).

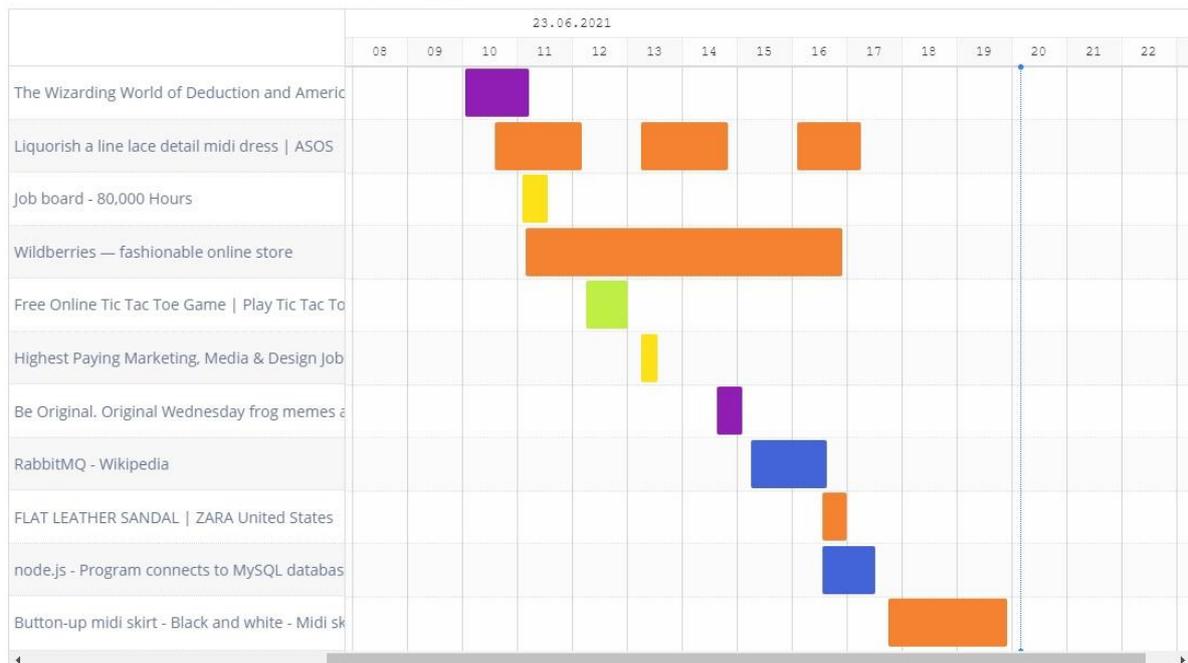


Рис.50. Таймлайн открытых вкладок

## ЗАКЛЮЧЕНИЕ

В ходе работы было разработано ПО для мониторинга и анализа активности пользователя на персональном компьютере. Была рассмотрена и проанализирована проблема неэффективности работы персонала, выделены преимущества внедрения ПО для мониторинга персонала, проведен сравнительный анализ существующих систем. В ходе анализа было выявлено три недостатка рассмотренных приложений.

Затем была проведена разработка и анализ требований к системе. Была сформулирована цель разработки, выявлены функции системы, ее окружение и поведение. Результатами стали диаграммы вариантов использования и потоков данных. На этапе проектирования была разработана архитектура исполнения система, представленная диаграммой развертывания. Также были спроектированы базы данных системы.

Далее были выбраны инструменты разработки, приняты решения о реализации конкретных блоков и произведена реализация основной части программного обеспечения.

Для обучения классификаторов был создан датасет, содержащий текст веб-страниц, разбитый на пять блоков по типу контента. На основании проведенных экспериментов в качестве классификаторов были выбраны один-против-всех для текста, названий и meta keywords и нейронная сеть для заголовков и meta description. Далее все пять классификаторов были обучены и протестированы. На основании полученных результатов, был сделан вывод о целесообразности учета полученных значений точности в качестве коэффициентов доверия. Была разработана формула для расчета оценки категории. Итоговая точность классификации, определенная экспериментально после применения всех преобразований и формулы - 87.4%.

Таким образом, была разработана система, которая:

- мониторит действия пользователей на персональных компьютерах посредством получения информации об открытых приложениях и вкладках браузера;
- при анализе веб-ресурсов классифицирует веб-страницы по тематикам с помощью машинного обучения с точностью 87.4%;
- определяет продолжительность пребывания в каждом приложении и на каждой странице;
- позволяет задавать “продуктивные” и “непродуктивные” тематики веб-ресурсов для конкретных рабочих групп;
- выводит результаты в виде графиков и тайм-лайнов;
- предоставляет информацию, которая помогает более эффективно принимать управленческие решения.

## СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1. Британские учёные доказали, что работать больше трёх часов в день — бессмысленно // Life URL: <https://life.ru/p/1049413> (дата обращения: 02.06.2021)
2. Какой процент рабочего времени вы реально работаете, а не бездельничаете? // Хабр URL: <https://habr.com/ru/post/86409/> (дата обращения: 02.06.2021)
3. Один день в поиске // Исследования Яндекса URL: <https://yandex.ru/company/researches/2017/oneday> (дата обращения: 02.06.2021)
4. Бесконечные чаепития // Зарплата.ру URL: <https://tyumen.zarplata.ru/articles/50610591> (дата обращения: 02.06.2021)
5. Статистика Интернета 2020: сайты, домены, хостинг, трафик // Студия Denisov URL: <https://sdvv.ru/articles/elektronnaya-kommertsiya/statistika-interneta-2020-sayty-domeny-khosting-trafik/> (дата обращения: 02.06.2021)