**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ**

Федеральное государственное бюджетное образовательное учреждение

высшего образования

**«КУБАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»**

**(ФГБОУ ВО «КубГУ»)**

**Факультет компьютерных технологий и прикладной математики**

**Отчет о выполнении лабораторной работы №2**

**по дисциплине**

**«Технологии проектирования программного обеспечения»**

Направленность Информационные технологии

Работу выполнил студент группы 4ПМ/2 И. А. Молчанов

(подпись)

Работу проверил доц. каф. ИТ, к.т.н., доц. А.Н. Полетайкин

(подпись)

Краснодар

2023

**ВВЕДЕНИЕ**

**Тема**: Анализ существующих компьютерных разработок.

**Цель работы**: ознакомление с существующими разработками подобных программных решений по выбранной̆ теме, приобретение навыков анализа существующих компьютерных разработок.

Для выполнения данный работы были поставлены следующие **задачи**:

1. Выполнить системное описание существующих подобных программных систем (не менее двух), которые могут быть применены к данному объекту управления; выделить основные преимущества и недостатки представленных систем.
2. Выполнить сравнительную характеристику описанных систем. Результаты сравнительного анализа представить в табличной̆ форме. Набор основных показателей̆ для сравнения:

* назначение системы;
* гибкость системы;
* защищенность системы;
* живучесть системы;
* надежность системы;
* открытость системы;
* мобильность;
* удобство пользовательского интерфейса системы;
* стоимость системы (в том числе затраты на тех. поддержку);
* эргономичность.

1. Сделать вывод о возможности или невозможности использования этих систем на выбранном объекте информатизации.

Собственная тема: Анализ существующих компьютерных разработок для анализа новостных сообщений.

* 1. **Описание ранее разработанных систем (подсистем)**
  2. **RapidMiner**

RapidMiner – это инструмент для анализа данных с графическим интерфейсом, который также может быть использован для анализа новостных сообщений. Он предоставляет набор операторов и функциональности, которые позволяют создавать конвейеры обработки данных для классификации новостей и извлечения полезной информации из текстовых данных.

В RapidMiner есть возможность создать процесс анализа данных, состоящий из нескольких этапов. В контексте анализа новостей, первым шагом может быть импорт данных из источника новостных статей, такого как база данных или файл. RapidMiner поддерживает различные форматы данных, включая CSV, Excel, JSON и другие.

После импорта данных можно применить операторы предварительной обработки текста для очистки и подготовки новостных статей для дальнейшего анализа. Например, реализовано использование операторов для удаления стоп-слов, токенизации (разделения текста на отдельные слова или токены), стемминга (приведение слов к их основной форме) и удаления ненужных символов или пунктуации.

Затем можно использовать операторы (Рис 1.1) для извлечения признаков из текста новостей. Это может включать операторы для извлечения ключевых слов, фраз, н-грамм (последовательностей из нескольких слов) или других характеристик, которые могут быть полезны для классификации или анализа новостей.

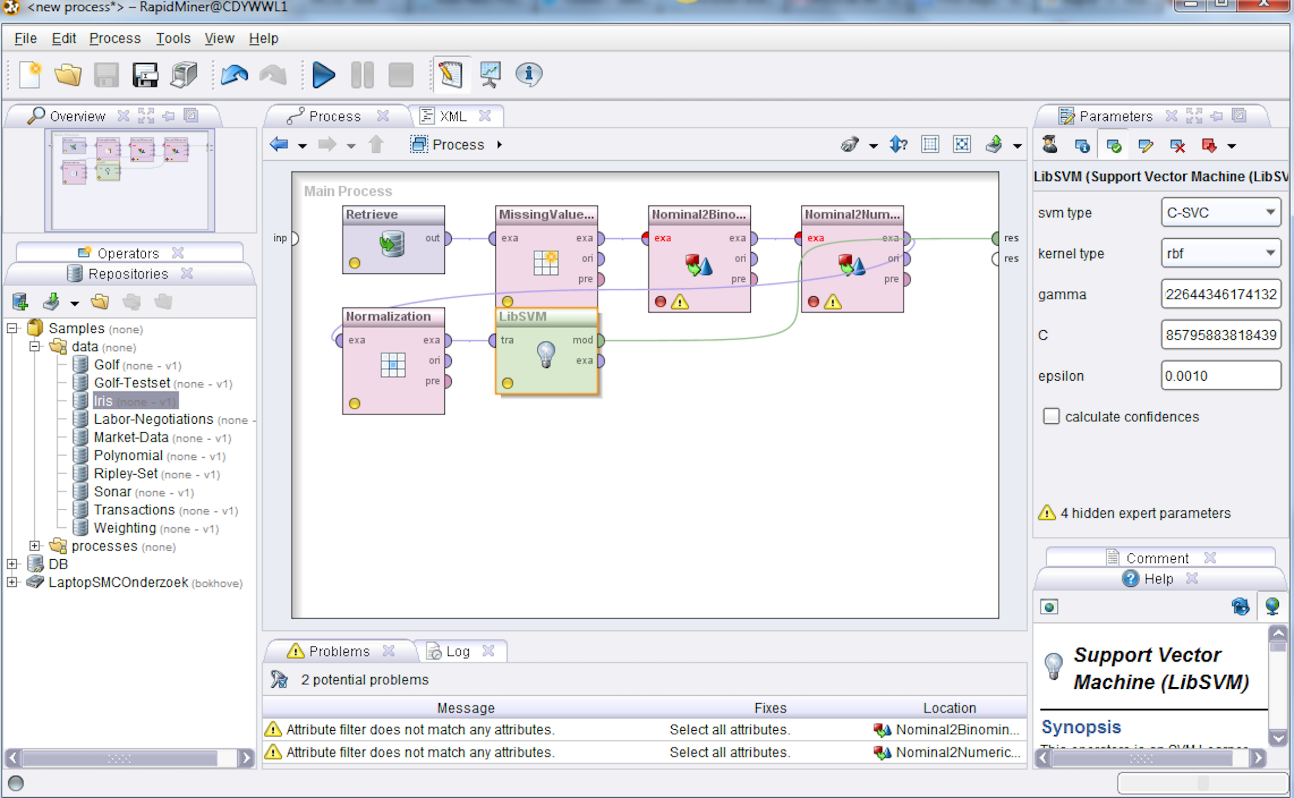


Рис 1.1 – Настройка операторов

После этого, если необходимо, можно применить алгоритмы классификации, доступные в RapidMiner, для классификации новостей. Это может быть метод наивного Байеса, машинное обучение на основе опорных векторов (SVM), решающие деревья или другие алгоритмы, которые предоставляются в RapidMiner. Доступна настройка параметров алгоритма и оценка его производительность с помощью операторов валидации и оценки модели.

Кроме классификации, RapidMiner также предоставляет возможности для анализа тональности текста, выявления тематик, кластеризации новостей и других видов анализа. Есть возможность использовать соответствующие операторы и методы, чтобы получить дополнительную информацию о ваших новостных данных.

Главное преимущество RapidMiner заключается в его графическом интерфейсе, который позволяет пользователю визуально создавать и настраивать процессы анализа данных без необходимости написания кода. Однако, RapidMiner также поддерживает использование скриптов на языке R или Python, если требуется более гибкий и расширенный анализ новостей. Интерфейс RapidMiner выглядит следующим образом (Рис. 1.2):

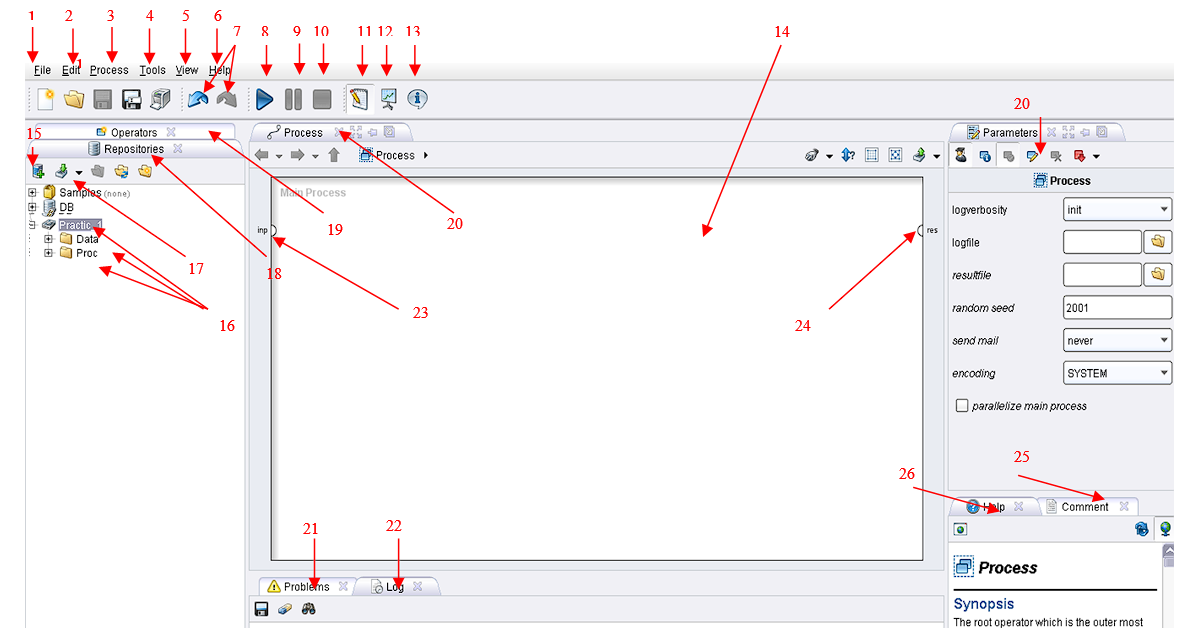


Рис 1.2 – Интерфейс RapidMiner

1. Пункт меню файл;
2. пункт меню вставка;
3. пункт меню процессы;
4. пункт меню инструменты;
5. пункт меню вид;
6. пункт меню помощь;
7. отмена и возврат действия;
8. запуск процесса;
9. кнопка приостановления процесса;
10. кнопка остановки процесса;
11. кнопка перехода на рабочий лист создания процесса;
12. кнопка перехода на рабочий лист результатов процесса;
13. кнопка перехода на приветственную страницу RapidMiner;
14. рабочее поле для создания процесса или модели;
15. кнопка для создания нового репозитория или загрузки уже созданного;
16. созданные папки для хранения данных (Data) и моделей, процессов (Proc) в репозитории (Practic\_1);
17. кнопка для импорта данных;
18. панель репозитория;
19. панель операторов;
20. панель параметры (меню будет видоизменятся в зависимости от активного элемента);
21. панель конфликтов. В нем будет появляться информация о возникших проблемах при выполнении процесса;
22. журнал загрузок;
23. узел входа данных (начало процесса);
24. узел выхода данных (завершение процесса);
25. меню комментариев;
26. меню помощи.

RapidMiner имеет множество графических представлений результата анализа текста (Рис 1.3-1.5):

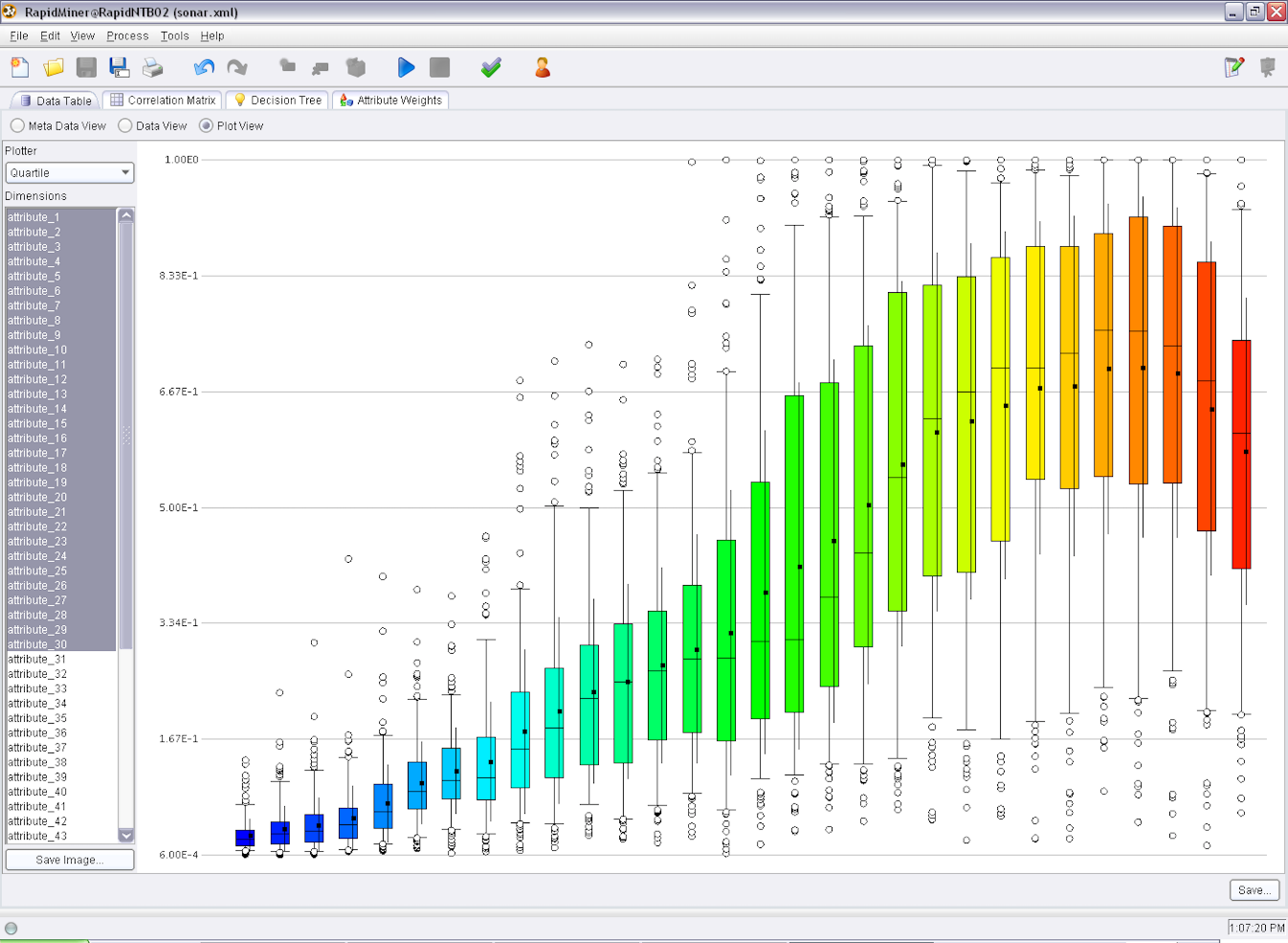


Рис 1.3 – Представление результата

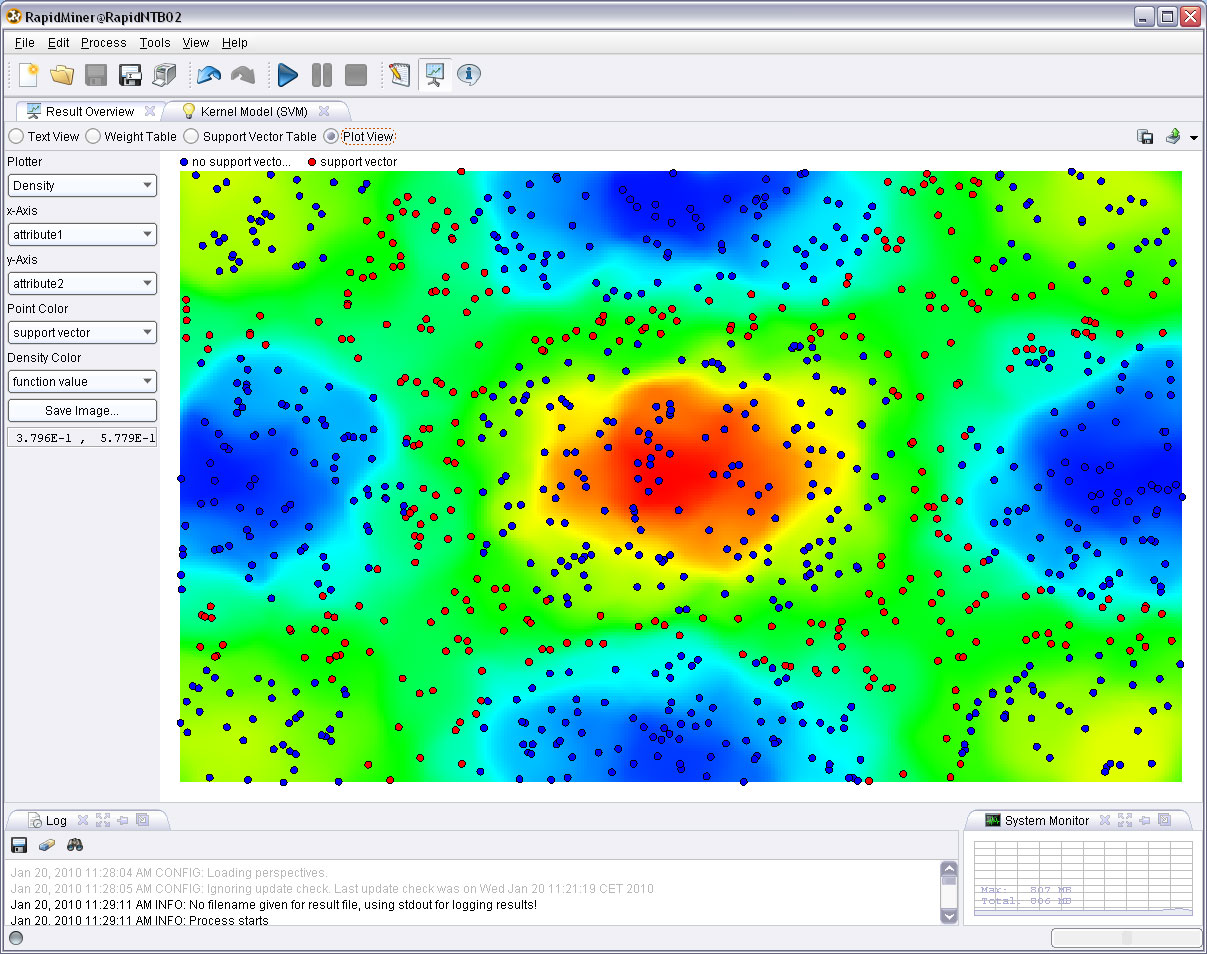


Рис 1.4 – Представление результата

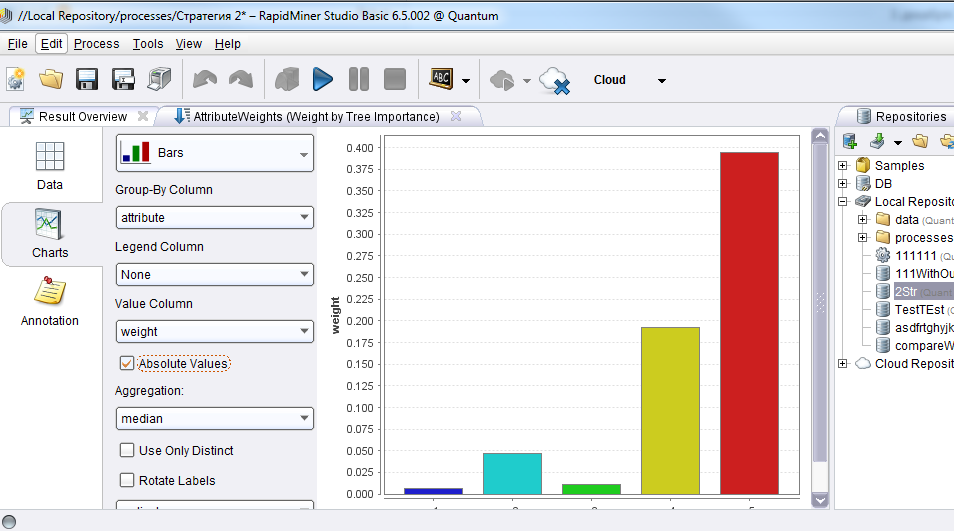


Рис 1.5 – Представление результата

В целом, RapidMiner предоставляет широкие возможности для анализа новостных сообщений и может быть полезным инструментом для создания и применения моделей классификации, извлечения признаков и других видов анализа в контексте новостных данных.

**1.2 MonkeyLearn**

MonkeyLearn – это платформа для автоматического обучения и классификации текста. Она предоставляет различные инструменты и API, которые позволяют пользователям создавать и развертывать модели машинного обучения для обработки текстовых данных, включая анализ новостных сообщений. Веб-интерфейс приложения MonkeyLearn (Рис 1.2.1):

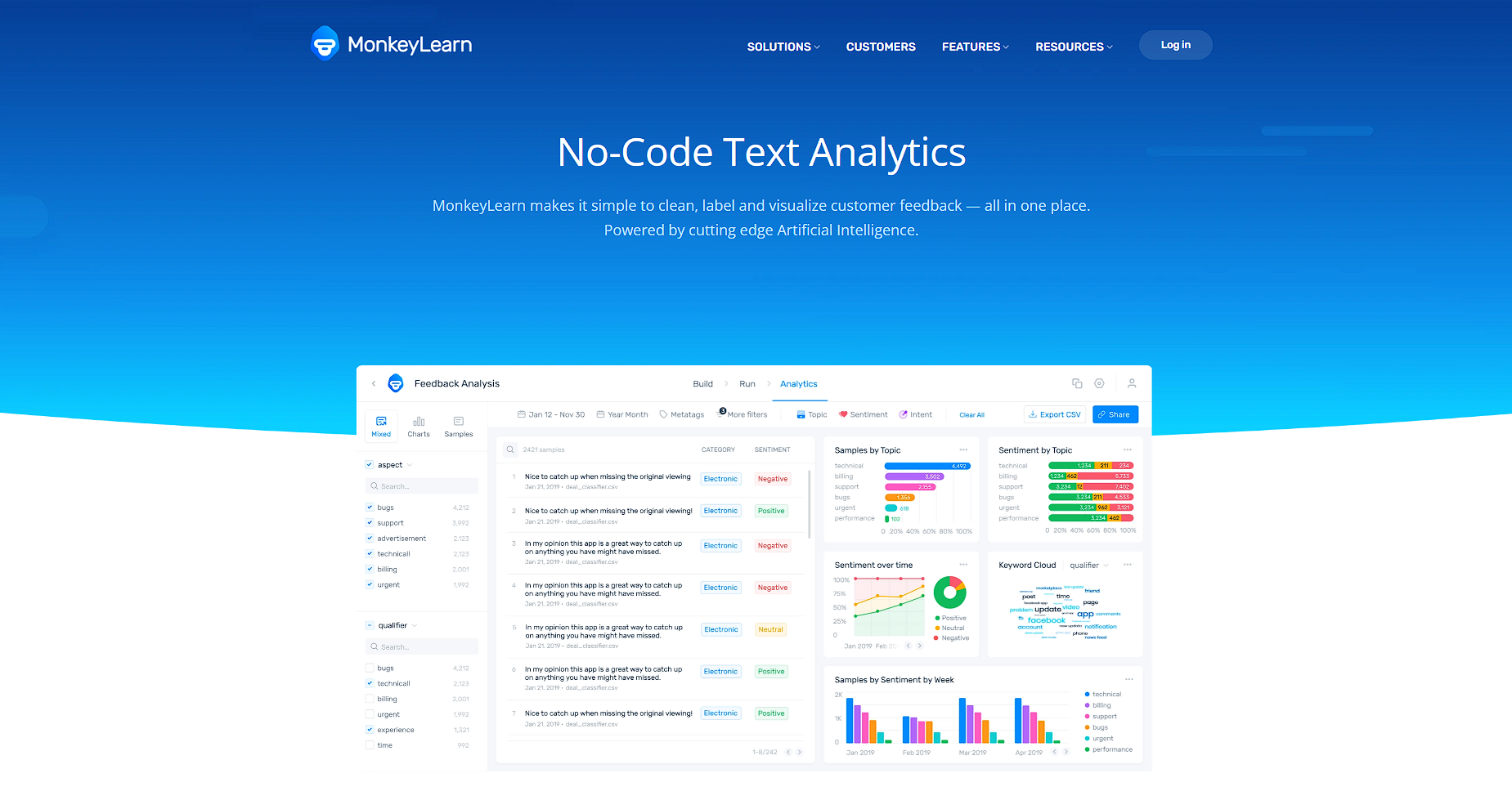


Рис 1.2.1 – Веб-интерфейс приложения MonkeyLearn

Основными функциональностями MonkeyLearn являются классификация текста, извлечение информации и обработка тональности. Эти функции позволяют обрабатывать текстовые данные, выделять ключевую информацию и анализировать эмоциональную окраску текста.

Для использования MonkeyLearn, вы можете создать модель классификации, которая будет принимать текстовые данные и присваивать им соответствующие метки или категории. В случае анализа новостей, вы можете создать модель, которая будет классифицировать новости по различным тематикам, например, спорт, финансы, политика и т.д. Для создания такой модели вам потребуются размеченные данные, то есть набор текстовых данных с уже известными метками категорий.

MonkeyLearn также предоставляет готовые модели и классификаторы, которые можно использовать без необходимости создания собственной модели. Вы можете использовать предварительно обученные модели, которые уже обучены на больших объемах данных и готовы к использованию в анализе новостей.

Помимо классификации, MonkeyLearn также предоставляет возможности для извлечения информации из текста. Это может включать извлечение ключевых фраз, сущностей, дат, местоположений и других типов информации, которая может быть полезна при анализе новостей.

Еще одна важная функция MonkeyLearn – это анализ тональности текста. С помощью этой функции вы можете определить эмоциональную окраску текста, например, положительную, отрицательную или нейтральную, что может быть полезно при анализе тональности новостных сообщений.

На сайте реализовано графическое представление результатов (Рис 1.2.2):

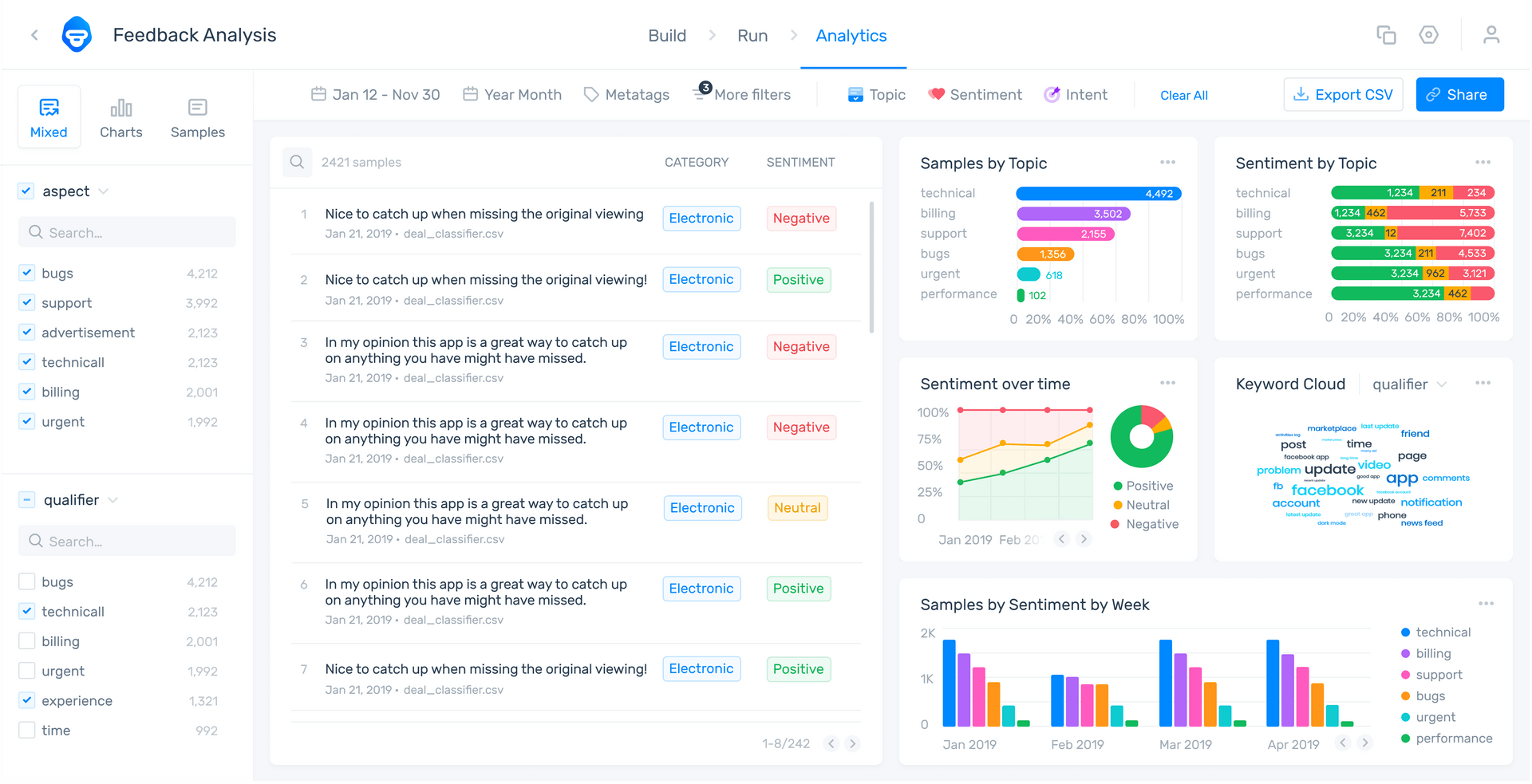


Рис. 1.2.2 – Графическое представление результатов анализа

MonkeyLearn предоставляет простой и интуитивно понятный интерфейс, который позволяет пользователям создавать и настраивать модели, загружать данные, обучать модели и получать результаты анализа. Кроме того, MonkeyLearn предлагает API, который позволяет интегрировать его функциональность в собственные приложения и системы.

* 1. **Сравнение систем**

Сравнение описанных ранее компьютерных разработок по одинаковым критериям представлено в таблице 1:

|  |  |  |
| --- | --- | --- |
| **Параметр** | **RapidMiner** | **MonkeyLearn** |
| Назначение | Инструмент для анализа данных и создания конвейеров обработки данных, включая анализ новостных сообщений. | Платформа для автоматического обучения и классификации текста, включая анализ новостных сообщений. |
| Гибкость | Предоставляет гибкие возможности для создания и настройки процессов анализа данных через графический интерфейс и скрипты на языке R/Python. | Предоставляет готовые модели и классификаторы, а также API для интеграции функциональности в приложения, с ограниченной возможностью настраивать модели. |
| Функ-ционал | Поддерживает различные операторы для предварительной обработки текста, извлечения признаков, классификации, анализа тональности и других видов анализа данных. | Предоставляет функционал для классификации текста, извлечения информации и анализа тональности, с возможностью использования предварительно обученных моделей или создания собственных. |
| Стоимость | Существует бесплатная версия с ограниченным функционалом. Платные планы доступны с различными функциями и уровнями поддержки. | Предлагает платные планы с различными функциями и уровнями использования. Бесплатный доступ ограничен по функциональности и количеству запросов. |
| Безопас-ность | Предоставляет возможность управления доступом и шифрования данных, а также соответствует нормам безопасности данных. | Обеспечивает безопасность данных с помощью шифрования и других мер безопасности, соответствует нормам GDPR и имеет меры контроля доступа. |
| Дизайн | Имеет графический интерфейс, который позволяет визуально создавать и настраивать процессы анализа данных без необходимости написания кода. | Имеет интуитивно понятный интерфейс, который удобен для использования и конфигурации моделей, а также предоставляет API для интеграции. |
| Пользовате-льский интерфейс | Имеет удобный и интуитивно понятный графический интерфейс, который облегчает создание и настройку процессов анализа данных. | Имеет простой и понятный интерфейс, который позволяет пользователям легко загружать данные, создавать и применять модели классификации. |
| Техничес-кая поддержка | Предлагает различные уровни технической поддержки для платных пользователей, включая документацию, форумы и обратную связь. | Предоставляет техническую поддержку в зависимости от выбранного плана, включая документацию, API-референции и техническую поддержку по электронной почте. |
| Системные требования | Работает на различных платформах (Windows, macOS, Linux) и требует установки на локальную машину или сервер. | Относительно менее требовательный продукт за счёт универсальной и удобной веб-версии. |

Таблица 1 – Сравнение «RapidMiner» и «MonkeyLearn»

* + 1. **Заключение о возможности применения систем к выбранному объекту**

**Вывод:** проделав лабораторную работу, я провел сравнительный анализ двух систем для ведения интернет-магазина. Обе системы, RapidMiner и MonkeyLearn, имеют функционал для анализа текстовых данных, включая анализ новостных сообщений. Однако, они имеют различия в терминах гибкости, функционала, стоимости, безопасности, дизайна, пользовательского интерфейса, технической поддержки и системных требований.

RapidMiner предоставляет гибкие возможности для создания и настройки процессов анализа данных через графический интерфейс и скрипты на языке R/Python. Он поддерживает различные операторы для предварительной обработки текста, извлечения признаков, классификации и анализа тональности. RapidMiner также предлагает различные уровни технической поддержки для платных пользователей. Однако, стоимость RapidMiner может быть выше, особенно для полного функционала и поддержки.

MonkeyLearn, с другой стороны, предлагает готовые модели и классификаторы для анализа текста, включая анализ новостных сообщений. Он также предоставляет API для интеграции функциональности в приложения. MonkeyLearn имеет простой и понятный интерфейс, что облегчает загрузку данных и создание моделей классификации. Однако, у MonkeyLearn ограниченные возможности настройки моделей по сравнению с RapidMiner.

Поэтому, при выборе между RapidMiner и MonkeyLearn для анализа новостных сообщений, следует учитывать следующие аспекты:

1. Если требуется высокая гибкость и возможность полной настройки процессов анализа данных, а также более широкий функционал, то RapidMiner может быть предпочтительным выбором. Он позволяет создавать сложные процессы обработки данных и имеет более мощные инструменты для работы с текстовыми данными.
2. Если важна простота использования и быстрая интеграция, а также если нужны готовые модели классификации текста, то MonkeyLearn может быть более подходящим вариантом. Он предоставляет простой интерфейс и предварительно обученные модели, которые можно быстро применить к анализу новостных сообщений.

Оба инструмента обладают надежной безопасностью и поддерживают меры контроля доступа. Однако, стоит учитывать стоимость каждой системы и выбрать соответствующий план, основываясь на конкретных потребностях и бюджете.

Также необходимо учесть системные требования каждой системы. RapidMiner требует установки на локальную машину или сервер, в то время как MonkeyLearn является платформой, доступной через веб-интерфейс или API.

В целом, обе системы предлагают функциональность для анализа новостных сообщений, но различаются в гибкости, возможностях настройки, стоимости и пользовательском опыте. Выбор между RapidMiner и MonkeyLearn зависит от конкретных требований, предпочтений и доступного бюджета.