Бегерський Михайло Васильович
(*прізвище, ім'я, по батькові студента*)

АНОТАЦІЯ

Дана магістерська дисертація присвячена розробці методу автоматизованої класифікації текстових даних на основі гібридних моделей.

Даний проект складається з двох основних частин: ресурсу для збору початкових даних та компоненту, що здійснює тренування та запуск моделей на отриманих даних. Частина для агрегації вхідних даних являє собою вебдодаток, що використовується для проходження опитування і формуванні результуючої таблиці на основі заповних форм. В якості проміжного етапу здійснюється підготовка та попередня обробка даних для подальшого використання. Підсистема класифікації в свою чергу поділена на два підмодуля: безпосередня імплементація розробленого методу класифікації та використання побудованої моделі для прогнозування зміни досліджуваної величини.

В рамках магістерської дисертації проведено аналіз існуючих систем для збору даних та розробка власної системи на основі адаптації готових рішень під вимоги досліджуваної області. Було здійснено дослідження існуючих алгоритмів для класифікації текстових даних та алгоритмів і бібліотек, що використовуються для побудових моделей прогнозування. Проведено оцінку предметної області, сформовано функціональні та нефункціональні вимоги до програмного забезпечення, а також проаналізовано перспективи виходу на ринок та запуску даного проекту в комерційних цілях.

У даній магістерській дисертації розроблено: архітектуру веб-ресурсу, інтерфейс для збору початкових даних, компнонент для обробки та трансформації даних, алгоритм для побудови прогностичної моделі та утиліту командного рядка для прикладного запуску моделі в якості інструменту прогнозування зміни цільової величини вхідних даних. Виконаний порівняльний аналіз з уже існуючими рішеннями та перевірка коректності роботи на основі порівняння відхилення результуючих показників з еталонними. Дана система готова розгортання, використання та інтеграції з іншими рішеннями, а також до впровадження в якості самостійного проекту на ринок, націленого на комерціалізацію продукту.

Misha Behersky
(*student first name, last name*)

ABSTRACT

This Master's dissertation is about creating a method for automatic text data classification based on blender models.

The project consists of two main parts: data collection module and a compoment responsible for training and launching a model on the input data. Subcomponent for input data aggregation is a web-application that is used by target users to take a survey and then to form a resulting table based on an information filled. As a part of main pipeline data transformation and preprocessing takes place. Classification system by its own consists of two connected modules: implementation of a classification method and application for prediction using the model built.

In the scope of dissertation analysis of current system for data mining was made. Also new project based on requirements from domain field using solutions from existing alternatives was created. Research on existing algorithms of text data classification and comparison of libraries was conducted. The research in the domain field was performed, functional and non-functional requirements for the software were generated. Possibilities for commercial launching of the project were considered and corresponding breakdown took place.

Within the Master's dissertation following components were created: architecture of a web-resource, user interface for collecting initial data, component for data transformation and preprocessing, algorithm for predictive model creation, command line utility for launching a model built on the dataset in order to predict target value. Developed solution was compared to competitors and final measurements about system's correctness was made based on reference data. System created is ready for deployment, direct usage and integration with existing solutions as well as moving forward to the market targeting commercialization of a product.

Студент(ка) _____ _____
(*підпис*) (*прізвище, ім'я, по батькові*)

Науковий керівник _____ _____
(*підпис*) (*посада, вчене звання, науковий ступінь, прізвище, ініціали*)