**BC/NW 2022 № 1 (39):9.2**

**ИНФОРМАЦИОННАЯ СИСТЕМА АВТОМАТИЧЕСКОЙ РУБРКАЦИИ НОВОСТНЫХ ТЕКСТОВ С ПРИМЕНЕНИЕМ АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ**

**Челышев Э. А., Раскатова М.В.**

В последние годы все более возрастает объем производимых человечеством данных. Так, например, общемировой объем данных 2018 году составлял 33 зеттабайтов, а к 2025 году, согласно прогнозам, вырастет до 175 зеттабайтов [1]. Этот факт объясняет все более возрастающую потребность в использовании средств автоматической обработки информации.

Целью данной работы является разработка ориентированной на конечного пользователя информационной системы автоматической рубрикации русскоязычных новостных текстов с использованием алгоритмов машинного обучения.

В исследовании был использован размеченный набор данных [2], из которого были выделены новостные статьи, относящиеся к девяти рубрикам. В ходе работы были обучены четыре классификатора: на основе наивного байесовского классификатора, логистической регрессии, случайного леса решающих деревьев и искусственной нейронной сети. Обобщающая способность каждого из классификаторов была оценена с использованием метрик , и -меры [3]. Для возможности удобного взаимодействия пользователя с системой был разработан веб-сайт с использованием фреймворка языка Python Django.

В данном докладе рассматривается информационная система автоматической рубрикации новостных текстов на основе классификатора, продемонстрировавшего наилучшую обобщающую способность, а также описывается пользовательский интерфейс и функционирование входящего в состав информационной системы веб-сайта.

**Литература**

1. **Reinsel, D.** The Digitalization of the World / D. Reinsel, J. Gantz, J. Rydning – 2018. – 28 с.: [Электронный ресурс]. – URL: <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>. (Дата обращения: 11.03.2021).
2. News dataset from Lenta.ru [Электронный ресурс] // Kaggle: Your Home for Data Science. URL: <https://www.kaggle.com/yutkin/corpus-of-russian-news-articles-from-lenta>. (Дата обращения 08.02.2021)
3. **Челышев, Э. А.** Разработка информационной системы для автоматической рубрикации новостных текстов / Э. А. Челышев, Ш. А. Оцоков, М. В. Раскатова // Международный журнал информационных технологий и энергоэффективности. – 2021. – Т. 6. – № 3(21). – С. 11-17.















