

УДК 004.4

Electronic service assistant for choosing professions (EsaCP), с применением ИИ Мейнов А.С. (ИТМО)

Научный руководитель – Директор центра сопровождения инклюзивного образования,
к.т.н., Гнездилова С.А. (ИТМО)

Введение. Каждый год количество абитуриентов, имеющих инвалидности или ограниченные возможности здоровья (ОВЗ), поступающих в ВУЗы, увеличивается. Для этой категории обучающихся крайне важно учитывать нозологию заболевания, физическую подготовку и индивидуальные возможности. Неправильный выбор специализации может привести к трудностям в обучении, что в свою очередь может отразиться на успеваемости, психологическом и физическом состоянии студента, и в конечном итоге привести к ухудшению здоровья.

Перед разработкой собственного электронного сервиса помощника я провел анализ рекомендаций выбора направления подготовки, с учетом нозологических групп абитуриента, указанных в методическом пособии "Поступаешь в Университет ИТМО? Сделай правильный выбор без ограничений".

Также важно помнить, что поддержка и адаптация образовательной среды для студентов с инвалидностями и ОВЗ играют ключевую роль в обеспечении равных возможностей для всех студентов.

Основная часть. Для реализации задачи по рекомендации выбора направления подготовки моя система основывается на модели дерева решений, которое охватывает несколько областей - классификацию и регрессию в машинном обучении. Алгоритм дерева решений будет использоваться для разделения функций набора данных, таких как вид и род деятельности. Затем устанавливаются параметры глубины дерева решений, чтобы снизить риск переобучения или слишком сложного дерева. Дерево решений состоит из разных узлов. Корневой узел — это начало дерева решений, которое обычно представляет собой весь набор данных в рамках машинного обучения. Листовые узлы являются конечной точкой ветви или конечным результатом серии решений. Дерево решений больше не будет разветвляться от конечного узла. В деревьях решений в машинном обучении характеристики данных являются внутренними узлами, а результатом является конечный узел.

Для того чтобы реализовать задачу по рекомендации выбора направления подготовки, моя система основывается на модели дерева решений, охватывающей несколько областей - классификацию и регрессию в машинном обучении. Алгоритм дерева решений используется для разделения функций набора данных, таких как вид и род деятельности. Затем устанавливаются параметры глубины дерева решений, чтобы снизить риск переобучения или формирования слишком сложного дерева. Структура дерева решений включает различные узлы:

- Корневой узел: Представляет начало дерева решений и обычно охватывает весь набор данных в рамках машинного обучения.

- Листовые узлы: Являются конечной точкой ветви или конечным результатом серии принимаемых решений. После достижения листового узла дерево решений больше не разветвляется.

- В деревьях решений характеристики данных выступают внутренними узлами, а результат представляет собой конечный узел.

В основе системы лежит матрица профессий, которая подробно изложена в методическом пособии ИТМО. [1, стр. 20-31] База данных обладает гибкостью, легко масштабируется благодаря удобной возможности добавления данных, что положительно сказывается на последующей работе. Исходный набор данных включает информацию, основанную на направлениях подготовки Университета ИТМО и одного из колледжей Санкт-

Петербурга. Данные хранятся в базе данных MariaDB на сервере, что способствует быстрому развитию системы и подключению новых образовательных учреждений. Кроме того, система предусматривает кэширование данных на сервере, где находится алгоритм, и обмен данными с помощью собственного RestAPI. Это дает возможность встраивать систему в любое мобильное приложение или на веб-сайт.

Абитуриент может не только получить результат с помощью прохождения теста, но и просто написать несколько предложений о себе, что он любит делать, какое его любимое увлечение. Это позволит более индивидуально подойти к каждому кто будет использовать данный сервис, ведь часто людей отпугивает тест, а написать или надиктовать голосом пару предложений иногда проще и удобнее.

Для этой функции используется библиотека машинного обучения FastText, которая разработанная и поддерживаемая командой Facebook AI Research (FAIR). Одной из ключевых особенностей FastText является возможность обучения моделей на больших текстах с высокой эффективностью. Библиотека также предоставляет быстрые методы для классификации текста и вычисления семантических аналогий. С FastText можно легко и быстро решать такие задачи, как определение языка, извлечение ключевых слов, суммаризация текстов, генерация текстов, анализ тональности, распознавание именованных сущностей и многое другое. FastText основан на идее, что слова можно представить как сумму субслов, то есть частей слов, которые имеют смысл. Например, слово "кот" состоит из трех субслов: "к", "о" и "т". Это позволяет FastText учитывать морфологию и семантику слов, а также обрабатывать неизвестные или редкие слова. FastText также использует линейные модели и иерархическую softmax для эффективной классификации текстов по разным категориям. Благодаря своей эффективности и возможностям работы с большими объемами текста, FastText нашел широкое применение в области обработки естественного языка, включая задачи классификации текста, поиска похожих документов, анализа тональности и других задач [5].

В данный момент существует простой веб-сайт с минимальными функциями и минимальным графическим интерфейсом для демонстрации работы алгоритма с API, который обеспечивает полный функционал.

При посещении сайта пользователь, на главной странице видит два варианта взаимодействия с сервисом, он может пройти тест или набрать текст вручную или голосом. Если выбрать прохождение теста, то перед пользователем появятся три вопроса, которые идут подряд. Из которых первые два обязательны для ответа, а на третий можно не отвечать. Пользователь может выбрать ответ, кликнув на его текст или поставив «галочку» рядом с ним. Затем в конце теста в специальном поле пользователь увидит результат пройденного теста

Если выбрать ввод текста, то пользователю станет доступно для ввода текстовое поле, где он может ввести текст удобным для него способом. Затем нажать на кнопку для вывода результата. Так же сейчас ведется разработка бота для Telegram, который имеет схожий функционал, как и на сайте.

Выводы. В данной работе представлено описание работы алгоритма машинного обучения для подбора направления подготовки для абитуриентов с инвалидностью и ограниченными возможностями здоровья. Алгоритм анализирует данные анкетирования и предлагает рекомендации по выбору специальности, учитывая особенности здоровья каждого абитуриента.

Список использованных источников:

1. С.А. Гнездилова, А.Ф. Джумагулова, Д.С. Минин, А.В. Мухаметшина // Поступаешь в Университет ИТМО? Сделай правильный выбор без ограничений // СПб. Университет ИТМО, Центр сопровождения инклюзивного образования – 2019.
2. Портал «www.mastersindatascience.org»
(<https://www.mastersindatascience.org/learning/machine-learning-algorithms/decision->

[tree/](#)) Просмотров: 02.02.2024

3. Building REST APIs with Flask: Create Python Web Services with MySQL // Kunal Relan // Apress; 1st ed. Edition – 2019
4. REST APIs with Django: Build powerful web APIs with Python and Django // William S. Vincent // Independently published – 2018
5. Портал «<https://fasttext.cc/>»
(https://fasttext.cc) Просмотров: 02.02.2024

Автор _____ Меинов А.С.

Научный руководитель _____ Гнездилова С.А.