

## АННОТАЦИЯ

Диссертационной работы на соискание ученой степени доктора философии  
Ph. D. по специальности 6D070400 - «Вычислительная техника и  
программное обеспечение» Таласбек Эсем Лесбекқызы  
«Определение склонности к профессии с помощью машинного обучения».

**Общая характеристика исследования.** Данная работа посвящена исследованию и разработке приложения, предлагающего рекомендации по выбору будущей профессии на основе личностных характеристик человека путем выявления профессиональных склонностей.

**Актуальность.** В настоящее время на Казахстанском рынке практически отсутствуют системы определения профессии. Современное общество предъявляет новые требования к производительности и профессионализму. Однако высокий уровень профессионализма предполагает полное раскрытие потенциала личности, что невозможно без учета личностных особенностей человека. Многие анкеты, проводимые организациями, недостаточно определяют и описывают тип человека для назначения, отбора персонала для определенных специальных программ и не дают надежного результата о том, справится ли данное лицо с определенными служебными обязанностями.

Консультации по вопросам карьеры призваны помочь людям научиться принимать мудрые и уверенные решения, связанные с карьерой. Это решение должно быть основано на правильном самопознании и внимательном рассмотрении множества альтернатив. Кроме того, люди должны быть удовлетворены своими решениями, успешно работать на выбранной ими работе и быть готовыми к изменениям в карьере или корректировкам в будущем. Личность - это совокупность характеристик и установок человека при решении различных социальных ситуаций в детских садах, школах, университетах, семьях, рабочих коллективах и т.д.

Люди склонны к предубеждениям и предрассудкам, которые могут повлиять на точность их суждений. Личность можно рассматривать как оценку в различных областях, таких как подбор персонала, выбор профессии, взаимоотношения и консультирование по вопросам здоровья. Личность очень сильно влияет на обучаемость человека. Например, в успеваемости мы можем видеть значительные различия между людьми, принадлежащими к экстравертам, и лицами, принадлежащими к интровертам. Одной из основных причин, по которым студенты бросают учебу в университетах, является низкая успеваемость, но личность также влияет на успеваемость на том же уровне, что и интеллектуальные способности, самооценка, мотивация и т. д. Исследования показывают, что личность может быть взята как эффективное средство измерения успеваемости, особенно на университетском уровне.

Прогнозирование типа личности, профессиональных наклонностей - одна из современных задач ученых. Рост использования социальных сетей, таких как

Twitter, Facebook, Instagram, привлек исследователей для автоматизированного прогнозирования личности и задач классификации. Ключевая теория этих исследовательских работ: BigFive(Большая пятерка), NEO-Personality-Inventory Revised, Ten Item Personality Inventory, Myers-Briggs Type Indicator (MBTI) и т. Д. Существующие работы в этой области основаны на применяемых алгоритмах контролируемого обучения. по эталонным наборам данных; однако основной проблемой для них являются данные, а точнее несбалансированные классы по признакам. Эта проблема усложняет задачу прогнозирования и классификации личности.

**Цель исследования.** Разработка приложения, которое определяет склонность человека к профессии на основе классификации личности, используя различные данные и применяя методы машинного обучения для достижения высокого уровня точности.

**Задачи исследования.** Следуя поставленной цели, в данной работе определены следующие задачи:

- изучение и анализ существующих типов личности и методы прогнозирования и классификации ее наклонностей
- изучение и анализ взаимосвязи между типами личности и профессией
- сбор и анализ данных из учетных записей социальных сетей для применения алгоритмов машинного обучения
- проведение экспериментов и внедрение модели для прогнозирования выявления профессиональных склонностей

**Объект исследования.** Основное внимание в исследовании уделяется автоматизированным методам определения профессии и типа личности.

**Методы исследования.** Поставленные задачи решены путем проведения теоретических и эмпирических исследований. В рамках исследования использовались концептуальные положения классических теорий и алгоритмов машинного обучения искусственного интеллекта, моделей глубокого обучения, исследований ведущих зарубежных и отечественных ученых в области рекомендательных систем, классификации личности, теории вероятностей, математической статистики, численного анализа, анализа данных в области информатики, психологии и образования.

**Научная новизна работы.** Новизна диссертации заключается в разработке автоматизированного метода определения профессии с учетом психологических особенностей человека. В исследовании были предложены результаты различных экспериментов, реализованных с использованием постов в Instagram и сочетания моделей рекуррентной нейронной сети (RNN) и сверточной нейронной сети (CNN).

**Научные утверждения подлежащие защите:**

- Методы и алгоритмы сбора данных;
- Методы и алгоритмы определения склонности к профессии;

- Модели для автоматической классификации личности на основе различных типов данных, собранных из Instagram;
- Эксперименты, результаты и обсуждение.

**Практическая значимость результатов исследования.** Практическая ценность диссертации заключается в улучшении услуг в сфере профориентации, академической успеваемости и возможности применения результатов исследования в различных системах рекомендаций для выпускников школ и университетов, что помогает улучшить системы, помогающие выявлять психотип и наклонности студентов, сотрудников, преступников и др.

**Публикации.** Опубликовано 6 работ, в том числе:

- 2 опубликованных статьи в журнале Q2 International Journal of Emerging Technologies in Learning (iJET), проиндексированном SCOPUS Percentile 66;
- 1 статья, соответствующая требованиям Высшей аттестационной комиссии МОН РК;
- 3 доклада в трудах международных конференций.

**Структура и объем диссертации.** Диссертация изложена на 82 страницах машинописного текста. Состоит из нормативных ссылок, определений, списка сокращений, введения, четырех основных глав, заключения, ссылок и приложения. Диссертация включает 26 таблиц, 36 рисунков. Список использованной литературы состоит из 104 наименований.

**В первой главе** представлена вводная часть, описаны проблемы и содержание исследовательской работы.

**Во второй главе** представлен обзор существующих работ, описаны методы профориентации и выявления личностных склонностей. Во-первых, глава охватывает теорию MBTI, которая классифицирует человека по шестнадцати типам личности. Теория включает восемь шкал, объединенных попарно: экстраверсия (E) - интроверсия (I), ощущение (S) - интуиция (N), мышление (T) - чувство (F), суждение (J) - восприятие (P). Функциональный портрет каждого типа личности описывается по порядку, в результате возникает склонность к определенной профессии.

Во-вторых, в главе рассматриваются подходы, которые используются в задачах прогнозирования и классификации типов личности. Глава делит методы на четыре категории: контролируемые, неконтролируемые, частично контролируемые и методы глубокого обучения. Описан литературный обзор современных исследовательских работ по каждой категории. Приведено сравнение и анализ результатов по работам каждой категории.

**Третья глава** описывает все прикладные эксперименты один за другим и анализирует их ограничения. Первый эксперимент в этой исследовательской работе был посвящен классификации личности с применением кластеризации k-means. Он состоит из трех частей: сбор данных, подготовка данных и реализация модели кластеризации k-means с настройкой гиперпараметров. Для сбора данных использовалась автоматизированная форма Google. При выборе трех кластеров

инерция составила 700. Свойство инерции представляет собой сумму квадратов расстояний от образцов до ближайшего центра кластера. Чем меньше инерционность, тем лучше работа. После настройки гиперпараметров и изменения количества кластеров до 16 наш параметр инерции был уменьшен до 107. В качестве следующего шага после шагов обучения и настройки гиперпараметров тестирование модели проводилось путем передачи некоторых случайных входных данных. В результате он определяет индекс кластера, которому он принадлежит.

Результаты эксперимента №2 показали, что люди, как правило, участвуют в опросах, если они ведутся на родном языке

В эксперименте №3 были реализованы модели Naive Bayes, XGBoost и Recurrent Neural Network. Сравнивались характеристики классического алгоритма и алгоритма глубокого обучения. Метрики оценки точности были рассчитаны для описания характеристик каждой модели.

Результаты эксперимента № 4 показывают, что CNN дает высокую производительность в изображениях, поскольку имеет фильтры и «детекторы состояния». Как и во всей работе, связанной с машинным обучением, ограничением этой работы является отсутствие данных. Еще одна вещь, на которую следует обратить внимание, это то, что CNN дает некоторые результаты, но не дает нам анализа характеристик и того, как они влияют на само предсказание.

Эксперимент №5 был проведен на Apache Spark на кластерах Google с использованием Google Colab и импорта библиотек машинного обучения. Для анализа скорости и эффективности полиномиальной наивной байесовской модели часть эксперимента №3 была переписана на PySpark с использованием парадигмы MapReduce, где метки являются ключами, а сообщения - значениями.

В эксперименте №6 использовался текст на казахском языке. В качестве выборки были взяты 450 бакалавров факультета компьютерных наук Университета Сулеймана Демиреля и предложена модель LSTM. Результаты показали эффективность предложенной модели.

**В четвертой главе** описана вся архитектура предложенного метода применения идентификации профессии. В приложении есть два режима для определения типа человека MBTI и его склонности к профессии. Первый режим основан на текстовых данных, а второй - на прогнозировании путем загрузки данных изображения. Скриншоты Android-приложения с пошаговой иллюстрацией представлены в тезисе.

**В пятой главе** представлены этапы тестирования, экспериментальные результаты и их сравнение между собой и существующими работами в области названия исследовательской работы.

**В заключении** обсуждаются анализ и результаты текущей работы, ее будущие направления.

Данная работа способствует идентификации профессиональных склонностей на основе личностных качеств, и результаты этой диссертации могут

быть применены в дальнейших исследовательских работах в области распознавания лиц, обработки естественного языка и компьютерных наук в области образования в целом.