

**Письменный отзыв официального рецензента
 На диссертационную работу Таласбек Эсем Лесбеккызы
 “Profession inclination identification using machine learning”, представленную на соискание ученой
 степени PhD по специальности 6D070400 - “Вычислительная техника и программное обеспечение”**

№ п/п	Критерии	Соответствие критериям	Обоснование позиции официального рецензента
1.	Тема диссертации (на дату ее утверждения) соответствует направлениям развития науки и/или государственным программам	1.1 Соответствие приоритетным направлениям развития науки и/или государственным программам 1) Диссертация выполнена в рамках проекта или целевой программы, финансируемого(ой) из государственного бюджета (указать название и номер проекта или программы) 2) Диссертация выполнена в рамках другой государственной программы (указать название программы) 3) Диссертация соответствует приоритетному направлению развития науки, утвержденному Высшей научно-технической комиссией при Правительстве Республики Казахстан (указать направление)	Работа посвящена исследованию возможностей выработке рекомендаций по профессиональной ориентации людей на основе их личностных характеристик. Работа лежит на стыке психологических способов изучения личности и методов обработки естественного языка. Таким образом, диссертация соответствует приоритетному направлению развития науки, утвержденному Высшей научно-технической комиссией при Правительстве Республики Казахстан, а именно, 3. Информационные, телекоммуникационные и космические технологии. 3.1.2. Речевые технологии и компьютерная лингвистика
2.	Важность для науки	Работа вносит/не вносит существенный вклад в науку, а ее важность хорошо раскрыта/не раскрыта	Тема исследования важна с точки зрения профессиональной ориентацией широкого круга людей, а предложенный метод, существенно снижающий трудоемкость выработки рекомендаций, вносит существенный вклад в теорию и практику профессионального

			ориентирования.
3.	Принцип самостоятельности	Уровень самостоятельности: 1) Высокий; 2) Средний; 3) Низкий; 4) Самостоятельности нет	Работа представляет собой законченное самостоятельное исследование, результаты которого подтверждены публикациями в журналах второго квартала соответствующей тематики.
4.	Принцип внутреннего единства	4.1 Обоснование актуальности диссертации: 1) Обоснована; 2) Частично обоснована; 3) Не обоснована. 4.2 Содержание диссертации отражает тему диссертации: 1) Отражает; 2) Частично отражает; 3) Не отражает 4.3. Цель и задачи соответствуют теме диссертации: 1) соответствуют; 2) частично соответствуют; 3) не соответствуют	Актуальность достаточно полно обоснована большим литературным обзором, который подтверждает значительный интерес научного сообщества в решении задачи классификации типов личности и профессиональной ориентации пользователей. Содержание работы полностью соответствует заявленной теме диссертации. Конечная цель исследования разработана приложением, которое определяет профессиональную склонность человека на основе классификации личности, используя текстовые и графические данные с применением методов машинного обучения для достижения высокого уровня точности. Задачи - изучить и проанализировать методы прогнозирования и классификации типов личности; - изучить и проанализировать взаимосвязь между типами личности и профессией; - собрать и проанализировать данные из учетных записей социальных сетей для применения

		<p>алгоритмов машинного обучения с определения психологического типа личности и ее профессиональных предпочтений; - разработать приложение для прогнозирования профессиональных склонностей.</p> <p>Поставленная цель и обозначенные задачи полностью соответствуют теме диссертации.</p>
	<p>4.4 Все разделы и положения диссертации логически взаимосвязаны: 1) полностью взаимосвязаны; 2) взаимосвязь частичная; 3) взаимосвязь отсутствует</p>	<p>Тезисы диссертации последовательно описывают ход исследований и полученные результаты. Диссертационная работа состоит из логически связанных литературного обзора, описания серии экспериментов и их результатов, описания архитектуры и интерфейса программного обеспечения, построенного по результатам экспериментов и результирующей части.</p>
	<p>4.5 Предложенные автором новые решения (принципы, методы) аргументированы и оценены по сравнению с известными решениями: 1) критический анализ есть; 2) анализ частичный; 3) анализ представляет собой не собственные мнения, а цитаты других авторов</p>	<p>Соискатель критически проанализировала большой объем литературы. Критический и сравнительный анализ присутствует также при описании результатов вычислительных экспериментов.</p>
<p>5. Принципы научной новизны</p>	<p>5.1 Научные результаты и положения являются новыми? 1) полностью новые; 2) частично новые (новыми являются 25-75%); 3) не новые (новыми являются менее 25%)</p>	<p>Соискатель провел серию экспериментов, которая в целом полна для имеющегося небольшого корпуса текстов и описывает применение как классических алгоритмов машинного обучения, так и нескольких глубоких моделей в сравнении. Эту часть исследования можно рассматривать и как новый результат на основе известных методов и определенных методический вклад в такого рода исследования с</p>

	<p>5.2 Выводы диссертации являются новыми? 1) полностью новые; 2) <u>частично новые (новыми являются 25-75%)</u>; 3) не новые (новыми являются менее 25%)</p>	<p>применением машинного обучения и методов обработки текстов.</p> <p>Результрирующая часть обладает новизной, которая подтверждает возможность применения методов обработки естественного языка в решении задачи определения типа личности и выработки рекомендаций по профессиональной ориентации пользователей.</p>
6.	<p>Обоснованность основных выводов</p>	<p>Предложено решение по определению психологического типа личности и ее профессиональных склонностей является частично новым. Ранее, по данным соискателя, в литературе подобное комплексное решение не упоминалось.</p>
7.	<p>Обоснованность основных выводов</p>	<p>Полученные и представленные в диссертационной работе научные результаты подтверждаются результатами вычислительных экспериментов.</p> <p>Достоверность результатов обеспечена использованием современных методов проведения исследований, что дает основание считать полученные результаты достаточно обоснованными и достоверными.</p>
	<p>Необходимо ответить на следующие вопросы по каждому положению в отгледельности: 7.1 Доказано ли положение? 1) доказано; 2) скорее доказано; 3) скорее не доказано; 4) не доказано</p>	<p>На защиту вынесены следующие положения: Методы сбора данных; Методы идентификации профессиональных предпочтений пользователей; Модели для автоматического определения типа личности на основе данных собранных в Инстаграмм; Результаты вычислительных экспериментов.</p>

		<p>Положения доказаны литературным анализом, вычислительными экспериментами и публикацией основных результатов исследования в журналах второго квартала (Q2), индексируемых в базе SCOPUS.</p>
	<p>7.2 Является ли тривиальным? 1) да; 2) нет</p>	<p>Вынесенные положения не являются тривиальными.</p>
	<p>7.3 Является ли новым ? 1) да; 2) нет</p>	<p>Являются в значительной степени новыми.</p>
	<p>7.4 Уровень для применения: 1) узкий; 2) средний; 3) широкий</p>	<p>Степень применимости результатов диссертации широкая, так как касается широкой массы людей желающих определить свой психологический портрет и профессиональные склонности.</p>
	<p>7.5 Доказано ли в статье? 1) да; 2) нет</p>	<p>выводы подтверждаются собственными статьями автора.</p>
<p>8.</p>	<p>Принцип достоверности Достоверность источников и предоставляемой информации</p>	
	<p>8.1 Выбор методологии - обоснован или методология достаточно подробно описана 1) да; 2) нет</p>	<p>Метод исследования включает машинное обучение, методы обработки естественного языка и комплекс вычислительных экспериментов. Метод исследования подробно описан и является общепринятым для подобного рода исследований</p>
	<p>8.2 Результаты диссертационной работы получены с использованием современных методов научных исследований и методик обработки и интерпретации данных с применением компьютерных технологий: 1) да; 2) нет</p>	<p>Методы вполне современные и, в целом, адекватны имеющемуся набору данных и поставленной цели исследования.</p>

	<p>8.3 Теоретические выводы, модели, выявленные взаимосвязи и закономерности доказаны и подтверждены экспериментальным исследованием (для направленный подготовки по педагогическим наукам результаты доказаны на основе педагогического эксперимента):</p> <p>1) да; 2) нет</p> <p>8.4 Важные утверждения подтверждены/частично подтверждены/не подтверждены ссылками на актуальную и достоверную научную литературу</p> <p>8.5 Использованные источники литературы <u>достаточно/не достаточно</u> для литературного обзора</p>	<p>Экспериментальные исследования подтверждают с достаточной степенью достоверности гипотезу исследования.</p> <p>Литературный обзор достаточно полно описывает теорию типов личности и методы машинного обучения, используемые для представления личностных характеристик, в том числе, на основе индивидуальных лингвистических характеристик. Тем самым основные положения диссертации соответствуют современному состоянию исследования по данной тематике.</p> <p>Использованный перечень из более чем 100 источников вполне достаточен.</p>
<p>9. Принцип практической ценности</p>	<p>9.1 Диссертация имеет теоретическое значение:</p> <p>1) да; 2) нет</p> <p>9.2 Диссертация имеет практическое значение и существует высокая вероятность применения полученных результатов на практике: 1) да; 2) нет</p> <p>9.3 Предложения для практики являются новыми? 1) полностью новые;</p>	<p>Диссертация имеет некоторое теоретическое значение, иллюстрируя, путем вычислительных экспериментов, связь типологии личности с характерными особенностями письма респондентов.</p> <p>Диссертация имеет практическое значение, поскольку предлагает более быстрый метод оценки личностных качеств на основе методов обработки естественного языка.</p> <p>На основе подтвержденных в ходе вычислительных экспериментов моделей машинного обучения, разработано Android</p>

	<p>2) частично новые (новыми являются 25-75%); 3) не новые (новыми являются менее 25%)</p>	<p>приложение для выявления склонностей пользователя на основе теории типов личности МБТИ. Приложение работает в двух режимах: на основе текста и на основе фотозображения пользователя.</p>
<p>10. Качество написания и оформления</p>	<p>Качество академического письма: 1) <u>высокое</u>; 2) среднее; 3) ниже среднего; 4) низкое.</p>	<p>Диссертация в целом написана грамотно, ясно, хорошо структурирована. Имеются некоторые незначительные недоработки, в целом, не снижающие качество изложения.</p>

Заключение:

Представленная работа посвящена исследованию возможностей рекомендаций по профессиональной ориентации людей на основе их личностных характеристик. Работа лежит на стыке психологических способов изучения личности и методов обработки естественного языка.

Автор, взяв за основу теории типов личности Майера-Бриггса, восходящую к идеям К. Юнга, обосновала возможность классификации личности на 16 типов с использованием методов классификации текстов и разработала соответствующее приложение. В целом работа представляет собой законченное исследование, результаты которого подтверждены публикациях в журналах второго квартала соответствующей тематики.

Тезисы диссертации последовательно описывают ход исследований и полученные результаты. Диссертационная работа состоит из литературного обзора, описания серии экспериментов и их результатов, описания архитектуры и интерфейса программного обеспечения, построенного по результатам экспериментов и результирующей части.

Литературный обзор достаточно полно описывает теорию типов личности и методы машинного обучения, используемые для предказания личностных характеристик, в том числе, на основе индивидуальных лингвистических характеристик. Соискатель исследовала большой объем литературы.

Серия экспериментов в целом полна для имеющегося небольшого корпуса текстов и описывает применение как классических алгоритмов машинного обучения, так и нескольких глубоких моделей. Эту часть исследования можно рассматривать и как определенный методический вклад в такого рода исследования с применением машинного обучения и методов обработки текстов.

Результирующая часть подводит итог проведенных экспериментов.

Несмотря на в целом достаточно полное описание и внутреннее единство работы, ее оценка будет неполна без учета замечаний и рекомендаций для дальнейших исследований.

Замечания.

Список аббревиатур не полон, например, полезно его дополнить аббревиатурами IE, S/N, F/T, J/P. Корпус текстов небольшой, поэтому применение технологии map reduce и конкретно пакета Apache Spark выглядит избыточным.

Таблица 5.2 требует более ясных пояснений полученных результатов. Не совсем понятно, как сочетаются опубликованные в литературе результаты и собственные результаты соискателя.

На рисунке 5.2. отображена, по существу, перетренировка конволюционной сети. В дальнейших экспериментах следует предпринять усилия по ее исключительно с применением регуляризаторов и/или dropout слоев.

Рекомендации для дальнейших исследований.

Для оценки предпочтений респондентов в области профессиональной ориентации уместно в дальнейших экспериментах использовать профориентационные тесты, с которыми проводить сравнение индивидуальных языковых моделей пользователей.

Полезным дополнением будет использование автоматических систем перевода, которые в настоящее время эффективно используются для анализа и сравнения полиязычных текстовых корпусов.

Анализ текстовой информации пользователей уместно провести и с использованием моделей внимания (Attention models), в том числе, предобученных, поскольку литературные источники свидетельствуют о хороших результатах, связанных с использованием данных моделей даже в задачах общего характера.

Данную работу в силу большого количества проведенных экспериментов с применением как классических моделей машинного обучения, так и некоторых моделей глубокого обучения целесообразно рассмотреть как отправную точку подобного рода исследований.

Выводы.

Приведенные замечания и рекомендации не умаляют полезности и научной обоснованности полученных в диссертации результатов. Работа, в целом, соответствует требованиям, предъявляемым к диссертационным работам на соискание ученой степени Ph.D. По результатам анализа диссертационной работы рекомендую Эсем Тапасбек степень доктора философии (Ph.D) по специальности 6D070400 – «Вычислительная техника и программное обеспечение»

Рецензент, Dr.sc.ing.,

профессор кафедры "Программная инженерия" Sabbaev University **Р.И. Мухамедиев**

