

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ ОБРАЗОВАТЕЛЬНОЕ
БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ
ФЕДЕРАЦИИ»

Кафедра «Информационных технологий и анализа больших данных»

Курсовая работа по дисциплине Технологии анализа данных и машинного
обучения на тему «Предварительный анализ данных в задачах расшифровки
текстов на древних языках»

Студентки группы ПИ20-2

Факультета Прикладной информатики

Зорихиной Майи Витальевны

(подпись)

Научный руководитель: Кочкаров А.А.

(подпись)

Москва 2023

ВВЕДЕНИЕ

Машинное обучение - это одна из самых перспективных областей в информационных технологиях. Оно может применяться в самых разных областях, включая расшифровку древних письмен. В прошлом исследователи тюркологии систематизировали письменные памятники, написанные на различных алфавитах, таких как рунические, арабский, греческий, кириллический алфавиты и китайские иероглифы. Современные технологии машинного обучения и компьютерной лингвистики могут помочь в расшифровке этих текстов, что позволит расширить наше знание о древних цивилизациях и их культуре.

Машинное позволяет компьютеру обучаться на основе опыта и данных, не требуя явной программной инструкции. Использование этой технологии для расшифровки текстов на древних языках может значительно упростить и ускорить процесс их изучения. Для применения методов машинного обучения в данной области необходимо иметь набор данных на древнем языке, а также набор соответствующих переводов на более современный язык. Эти данные могут быть получены путем археологических раскопок, из исторических архивов, а также из частных коллекций.

Одним из наиболее распространенных методов машинного обучения для расшифровки древних текстов является использование нейронных сетей. Нейронные сети могут обучаться распознаванию определенных шаблонов в данных и использоваться для расшифровки неизвестных слов и фраз в тексте на древнем языке.

Кроме того, машинное обучение может использоваться для создания программ, которые могут автоматически определять язык, на котором написан текст, и переводить его на другой язык. Это позволяет облегчить процесс перевода и сделать его более точным. Для проведения исследования по применению методов машинного обучения для расшифровки текстов на древних языках, необходимо иметь соответствующий набор данных. Этот набор данных должен включать в себя множество текстов на древних языках,

написанных различными письменными системами, сопровождаемых переводами на современные языки, и метаданными, такими как год создания текста, его автор и тематика. Изучение и подготовка набора данных для расшифровки текстов на древних языках может быть сложным и трудоемким процессом, так как источники на древних языках могут быть редкими и труднодоступными, а также могут содержать различные варианты написания слов и грамматических форм.

Однако следует отметить, что применение машинного обучения в этой области также имеет свои ограничения. Например, для эффективной работы алгоритмов необходимо иметь достаточно большой набор данных на древнем языке, что может быть сложно для редких языков или языков, которые уже вышли из употребления. Также возможны ошибки в распознавании символов, которые могут привести к неточному переводу текста.

В любом случае, использование методов машинного обучения для расшифровки древних текстов открывает новые возможности для исследования истории и культуры нашей цивилизации, позволяя получить новые знания и пересмотреть старые теории.

Научная новизна темы заключается в том, что данный подход позволяет автоматически распознавать древние символы, проводить анализ текстов и определять вероятный перевод слов, используя большой объем данных. Ранее эту задачу решали вручную ученые-эксперты, что требовало большого количества времени и не всегда давало точные результаты. Применение методов машинного обучения позволяет автоматизировать процесс и ускорить получение результатов. Более того, этот подход может использоваться для расшифровки новых текстов, которые еще не были исследованы, что открывает новые возможности для изучения древних культур и языков.

Тема, связанная с расшифровкой древних текстов при помощи методов машинного обучения, считается актуальной и интересной в наше время. Несмотря на то, что существует множество не расшифрованных древних

письменностей, применение современных технологий может помочь ученым в этой области получать более точные результаты в более короткие сроки.

Во-первых, многие древние письменности до сих пор не полностью расшифрованы, и мы не имеем доступа к многим знаниям, заключенным в этих текстах. Многие из них имеют большое историческое и культурное значение, и их расшифровка может помочь нам лучше понять историю и культуру наших предков.

Во-вторых, использование методов машинного обучения может помочь в расшифровке текстов быстрее и эффективнее, чем ручная работа ученых. Это может существенно ускорить процесс расшифровки и дать более точные результаты.

Наконец, данная тема является актуальной в контексте развития искусственного интеллекта и машинного обучения. Применение этих методов для расшифровки древних текстов может помочь улучшить алгоритмы и программы машинного обучения, что в свою очередь может привести к созданию более эффективных и точных инструментов для работы с данными в различных областях науки и технологии.

Расшифровка древних текстов имеет огромное историческое и культурное значение. Многие из этих текстов содержат множество сведений о жизни наших предков, и их расшифровка может помочь ученым получить новые знания о культуре, истории и традициях древних цивилизаций. Однако, следует заметить, что расшифровка древних текстов является сложным и трудоемким процессом. Для успешной работы с древними текстами необходимо применять современные методы машинного обучения и компьютерные технологии, а также обладать широкими знаниями в области истории, культуры и лингвистики. В своей работе я попытаюсь написать код на языке Python с использованием машинного обучения чтобы реализовать предобработку данных в задаче расшифровки текстов на древних языках.

