



Universitatea Tehnică a Moldovei

**ANALIZA INTELECTUALĂ DATELOR PRIVIND
REZULTATELE VINULUI INTERNATIONAL CONCURSURI**

**INTELLECTUAL ANALYSIS OF DATA ON THE RESULTS OF
INTERNATIONAL WINE COMPETITIONS**

Student: gr. TIA-201M,
Bodnariuc Zoia

Conducător: Cunev Veaceslav
lector universitar

Chișinău 2022


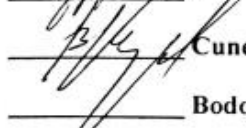
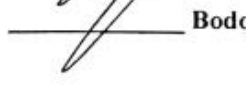
MINISTERUL EDUCAȚIEI ȘI CERCETĂRII AL REPUBLICII MOLDOVA
Universitatea Tehnică a Moldovei
Facultatea Calculatoare, Informatică și Microelectronică
Departamentul Ingineria Software și Automatică

Admis la susținere
Șef departament:
FIODOROV Ion dr., conf.univ.

„__” _____ 2022

ANALIZA INTELLECTUALĂ DATELOR PRIVIND
REZULTATELE VINULUI INTERNATIONAL
CONCURSURI

Proiect de master

Student:  Bodnariuc Zoia, TIA-201M
Coordonator:  Cunev Veaceslav, lect. univ.
Consultant:  Bodoga Cristina, asist.univ.

Chișinău 2022

АННОТАЦИЯ

Магистерская работа на тему: Интеллектуальный анализ данных по итогам международных винных конкурсов (по проекту «GUSTOS LIFE»).

Автор: Боднарюк Зоя, студентка группы ТИА-201М

Ключевые слова: анализ данных, Data Mining, машинное обучение, Python.

Работа состоит из введения, трех глав, заключения, списка использованной литературы и приложений.

Во введении раскрывается актуальность выбранной темы, описываются цель и задачи работы.

В первой главе изложены теоретико-методологические основы по выбору методов, инструментов и программного обеспечения для осуществления: сбор данных и создание проблемы, обработка данных, машинное обучение и интеллектуальный анализ данных, визуализация данных, генерация потенциальных бизнес-решений.

Вторая глава посвящена изучению и оценке критерий вина на конкурсах, структуры базы данных, выявление возможных гипотез для построения рекомендации по завоеванию лучших оценок на соревнованиях

В третьем главе включает в себя построение гипотез на основе данных по международным винным конкурсам, предоставленной базы данных в формате SQL компанией SRL «DEEPLACE», построение рекомендаций

В заключении обобщается проделанная работа и формулируются выводы по выявленным по результатам анализа данных, предложены рекомендации для потенциальных участников соревнований.

Работа содержит 51 страницы, 1 таблицы, 19 рисунков, списка использованных источников, 2 приложений.

REZUMATUL

Proiect de master pe tema: Analiza intelectuală datelor privind rezultatele vinului international concursuri (în cadrul proiectului „GUSTOS LIFE”).

Autor: Bodnariuc Zoia, student a gr. TIA-201M.

Cuvinte cheie: analiza datelor, DATA MINING, machine learning, PYTHON.

Lucrarea constă dintr-o introducere, trei capitole, o concluzie, o listă de referințe și aplicații.

Introducerea dezvăluie relevanța temei alese, descrie scopul și obiectivele lucrării.

Primul capitol conturează bazele teoretice și metodologice ale alegerii metodelor, instrumentelor și soft de implementare: colectarea datelor și crearea de probleme, prelucrarea datelor, învățarea automată și data mining, vizualizarea datelor, generarea de potențiale soluții de afaceri.

Al doilea capitol este dedicat studiului și evaluării criteriilor vinului la concursuri, structurii bazei de date, identificării posibilelor ipoteze pentru construirea de recomandări pentru obținerea celor mai bune note la concursuri.

Al treilea capitol cuprinde construirea de ipoteze pe baza datelor din competițiile internaționale de vinuri furnizate de baza de date în format SQL de către SRL „DEEPLACE”, construirea de recomandări.

În concluzie, se sintetizează munca depusă și se formulează concluzii asupra datelor identificate prin rezultatele analizei, se propun recomandări pentru potențialii participanți la concurs.

Lucrarea conține 51 de pagini, 1 tabel, 19 figuri, o listă de referințe, 2 aplicații.

ABSTRACT

Master's project on the topic: Data mining based on the results of international wine competitions (under the project "GUSTOS LIFE").

Author: Zoya Bodnariuc, student of the TIA-201M group

Keywords: data analysis, DATA MINING, machine learning, PYTHON.

The work consists of an introduction, three chapters, a conclusion, a list of references and applications.

The introduction reveals the relevance of the chosen topic, describes the purpose and objectives of the work.

The first chapter outlines the theoretical and methodological foundations for choosing methods, tools and software for implementation: data collection and problem creation, data processing, machine learning and data mining, data visualization, generation of potential business solutions.

The second chapter is devoted to the study and evaluation of wine criteria at competitions, the structure of the database, the identification of possible hypotheses for building recommendations for winning the best marks at competitions.

The third chapter includes the construction of hypotheses based on data from international wine competitions provided by the database in SQL format by SRL "DEEPLACE", the construction of recommendations

In conclusion, the work done is summarized and conclusions are formulated on the data identified by the results of the analysis, recommendations for potential participants in the competition are proposed.

The work contains 51 pages, 1 table, 19 figures, a list of references, 2 applications.

СОДЕРЖАНИЕ

ВВЕДЕНИЕ	8
1 ИНТЕЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ И ПРИМЕНЯЕМЫЕ ИНСТРУМЕНТЫ	10
1.1 Сбор данных и инструменты для создания проблем.....	11
1.2 Инструменты Data Science для обработки больших данных.....	12
1.3 Инструменты подготовки данных или обработки данных.....	13
1.4 Инструменты преобразования данных.....	15
1.5 Инструменты исследовательского анализа данных (EDA).....	16
1.6 Инструменты машинного обучения и принятия решений.....	17
1.7 Инструменты визуализации данных.....	19
2 АНАЛИЗ ПРЕДМЕТНОЙ ОБЛАСТИ, ВЫЯВЛЕНИЕ ВОЗМОЖНЫХ ГИПОТЕЗ	21
2.1 Критерии системы оценки вина на конкурсах.....	22
2.2 Структура базы данных Gustos.....	24
2.3 Выявление возможных гипотез.....	30
3 АНАЛИЗ ДАННЫХ ПО РЕЗУЛЬТАТАМ МЕЖДУНАРОДНЫХ КОНКУРСОВ ВИНА, ПРОВЕРКА ГИПОТЕЗ, ПОИСК ЗАРАНЕЕ НЕИЗВЕСТНЫХ ЗАКОНОМЕРНОСТЕЙ	32
3.1 Рейтинг соревнований, участников и цвета.....	33
3.2 Количество сахара среди участников с лучшими результатами.....	36
3.3 Количество алкоголя среди участников с лучшими результатами.....	39
ЗАКЛЮЧЕНИЕ	42
БИБЛИОТЕКА	43
ПРИЛОЖЕНИЕ А	44
ПРИЛОЖЕНИЕ В	47

ВВЕДЕНИЕ

Мы живем в веке информации. Трудно переоценить значение данных, которые мы непрерывно собираем в процессе нашей деятельности, в управлении бизнесом или производством, в банковском деле, в решении научных, инженерных и медицинских задач. Мощные компьютерные системы, хранящие и управляющие огромными базами данных, стали неотъемлемым атрибутом жизнедеятельности, как крупных корпораций, так и даже небольших компаний.

На протяжении последних десятилетий в информационной индустрии наблюдается устойчивый рост интереса к технологиям анализа данных и основанным на этих технологиях системам поддержки принятия решений. За всей областью закрепился, ставший уже привычным в англоязычной литературе, термин Data Mining или Knowledge Discovery. На русском языке употребляют также словосочетание "интеллектуальный анализ данных".

Популярность Data Mining сегодня можно сравнить с его популярностью пол столетия назад, на заре компьютерной эпохи. Тогда, правда, этот термин не был известен, но много говорили об искусственном интеллекте, о нейронных сетях и распознавании образов. Однако, за немногими исключениями, практическую реализацию теории пришлось отложить до тех пор, пока аппаратная и программная инфраструктура не развилась до современного уровня. И сегодня, по завершению пятидесятилетнего цикла развития, мы вновь обращаемся к решению задач анализа, обладая уже для этого мощными вычислительными системами и системами управления базами данных, развитой операционной и языковой средой.

Интеллектуальный анализ данных – позволяет провести наиболее полный и глубокий анализ проблемы, дает возможность обнаружить скрытые взаимосвязи, принять наиболее обоснованное решение. В процессе своей деятельности промышленные предприятия, корпорации, ведомственные структуры, органы государственной власти и местного самоуправления накопили большие объемы данных. Они хранят в себе большие потенциальные возможности по извлечению полезной аналитической информации, на основе которой можно выявлять скрытые тенденции, строить стратегию развития, находить новые решения.

Новыми компьютерными технологиями, образующими ИАД являются экспертные и интеллектуальные системы, методы искусственного интеллекта, базы знаний, базы данных, компьютерное моделирование, нейронные сети, нечеткие системы. Современные технологии ИАД

позволяют создавать новое знание, выявляя скрытые закономерности, прогнозируя будущее состояние систем. Основным методом моделирования социально-экономического является метод имитационного моделирования, дает возможность на модели проиграть различные стратегии развития, сравнить альтернативы, учесть влияние многих факторов, в том числе с элементами неопределенности.

Интеллектуальный анализ данных - это фильтрация большого количества исходных данных для получения полезной информации, что дает компаниям конкурентное преимущество. Эта информация состоит из значимых форм и тенденций, которые уже присутствуют в данных, но ранее были неочевидными.

Нейронные сети ориентируются на правила использования данных, которые основаны найденных связях или на образце набора данных. Как результат, программное обеспечение постоянно анализирует значения и сравнивает его с другими факторами, и он сравнивает эти факторы неоднократно, пока не найдет новый шаблон. Эти шаблоны известны как правила. Затем программное обеспечение ищет другие шаблоны, основанные на этих правилах, или посылает оповещение, когда значение триггера выпадает.

Целью данной магистерской работы является интеллектуальный анализ результатов международных винных конкурсов, выявление закономерностей (если таковые присутствуют), построение гипотез и проверка их. Изучение методов, инструментов и программного обеспечения для осуществления:

- сбор данных и создание проблемы;
- обработка данных (обработка данных или подготовка данных);
- машинное обучение и интеллектуальный анализ данных;
- визуализация данных;
- генерация потенциальных бизнес-решений.

Выбор оптимальных инструментов для анализа, выявление закономерностей или отрицание их, построение гипотез и проверка их, для дальнейшего увеличения призовых мест на международных конкурсах.

БИБЛИОТЕКА

1. Интеллектуальный-Анализ-Данных [прочитано 28 февраля 2022]. Доступно: <https://www.skachatreferat.ru/referaty/Интеллектуальный-Анализ-Данных/82757932>
1. Sangeeta Gulia, Top 10 Data Science Tools for 2022 [прочитано 28 февраля 2022]. Доступно: <https://www.techgeekbuzz.com/data-science-tools/>
2. Конкурсы и престижные премии лучшим винам, сомелье и виноделам года [прочитано 28 февраля 2022]. Доступно: <https://www.vins-france.ru/concours.htm>
3. Python и статистический вывод: часть 3 [прочитано 03 мая 2022]. Доступно: <https://habr.com/ru/post/556852/>
4. Сладость и сухость напитка [прочитано 03 мая 2022] Доступно: <https://ru.wikipedia.org/wiki>
5. Сколько вешать в граммах: все о сахаре в вине [прочитано 03 мая 2022] Доступно: <https://irk.winestate.ru/skolko-veshat-v-grammah-vse-o-sahare-v-vine>