WiDS '22 - '23 Final Documentation

Project 20 - Wine Quality Prediction

Ayushh Garg
Shlesh Gholap
Umang Agarwal

Team Member Name	Roll Number	Email-Id
Ishaan Manhar	210070033	ishaanmanhar2002@gmail.com
Joshitha Chowdhary	210050033	joshithachowdhary@gmail.com
Kartik Khandelwal	210040073	210040073@iitb.ac.in
Rudraksh Sangore	21D070060	rudraksh9sangore@gmail.com
Sandeep Kumar	22N0104	realsandeepkumar@gmail.com
Shatayu Ganvir	200110103	shatayuganvir@gmail.com
Siddharth Kaushik	210070086	sidu.kaushik@gmail.com

Introduction to Problem Statement

The entire overview of the project is available at:
https://lightning-bolt.notion.site/lightning-bolt/WiDS-Wine-Quality-Prediction-6ae2451f6d7544bfa168af089375c11

Existing Resources

Link to the resources page is:
https://lightning-bolt.notion.site/Resources-89617cbf0a604a2394bd9a1454fdbb7e

Proposed Solution

Solution made by one of the mentors is:
https://github.com/AgarwalUmang/Projects

Methodology & Progress (Mention the work done week-wise)

Week 1 - Basic Skills

- Revisiting Python programming language
- Getting familiar with Version Control Systems(VCS) such as Git, as well as GitHub.

Week 2 - Python Libraries

- Familiarizing and understanding relevant Python Libraries
- : Numpy, Pandas, Matplotlib, Seaborn (optional)

Week 3 - Introduction to Machine Learning

- Complete the first course on Machine Learning by Andrew NG on coursera as provided in the resources

Week 4 - Getting to know scikit-learn

- Watch the videos on scikit-learn as provided in the resources.
- Complete videos 1-5 and 14-17
- Play around with scikit-learn and familiarize yourself

Week 5 - Final Project Implementation

We head out to the final implementation of the WiDS Wine Quality Prediction. Project will be carried out in 2 parts.

First part has Exploratory Data Analysis and the next part will comprise Machine Learning.

Week 6 - Introduction to Neural Networks (optional)

Watch the playlist on Neural Networks by 3Blue1Brown and complete the remaining courses of Andrew NG's Machine Learning specialization. The link is in the resources section. Also go through the textbook provided on Neural Networks.

Above and Beyond

Read the research paper provided on Wine Quality Prediction (provided in the resources section) and implement the 1-D CNN model.

Results

Github repository containing the codes of all the mentees is mentioned below:
https://github.com/AgarwalUmang/WiDS-Wine-Quality-Prediction

Learning Value

Students learn from scratch. Basics of python is introduced through various tutorials.
Various python libraries useful for data science are learned. 0verview of libraries such as numpy, seaborn, matplotlib and pandas is given.
Andrew NG's machine learning course is completed and related theories are covered.
Scikit learn is extensively understood and implemented for the project,
Extensive data analysis is carried out by using the above libraries before going for basic ML.

This project is necessary for ML-enthusiasts. Foundations of ML are introduced to excite their interest and know more about this field in a basic sense.

Tech-stack Used

Python Libraries like numpy, pandas, matplotlib and seaborn.
Usage of scikit-learn for final project implementation.
Basics of pytorch/tensorflow is introduced.

Suggestions for others

The objective of the project was to provide students with basic ideas about data analysis and machine learning. Data cleaning and studying that data lays the foundation for complex machine learning projects. It is key to go through the fundamentals of machine learning to understand it in a better way.

Contribution by each Team Member

Each member individually implemented the project as mentioned in the work wise planner. Final code by them is uploaded to a common github repo and shared with the Analytics Club, IITB.

Members started with revision of Python and familiarizing themselves with various Python libraries necessary for data analysis and machine learning. Scikit-learn is introduced and Andrew NG's machine learning course on Coursera is completed. Everyone tried out a basic assignment for python revision and then moved on to the final assignment.
Final assignment consists of two parts: Data Analysis and basic Machine Learning. The members listed above successfully completed the final assignment.

References and Citations

Research Paper on Wine Quality Prediction:

https://s3.us-west-2.amazonaws.com/secure.notion-static.com/13bb49bf-d99a-4e00-ade5-2b5d9c5bd409/Wine_Qaulity.pdf?X-Amz-Algorithm=AWS4-HMAC-SHA256&X-Amz-Content-Sha256=UNSIGNED-PAYLOAD&X-Amz-Credential=AKIAT73L2G45EIPT3X45%2F20230131%2Fus-west-2%2Fs3%2Faws4_request&X-Amz-Date=20230131T043638Z&X-Amz-Expires=86400&X-Amz-Signature=42acceecec0516f4b56c7bbe0f6f9142f8fad7e0990c84c8ee34214205e25a46&X-Amz-SignedHeaders=host&response-content-disposition=filename%3D%22Wine%2520Qaulity.pdf%22&x-id=GetObject

Red Wine Quality Prediction project on Kaggle

https://www.kaggle.com/datasets/uciml/red-wine-quality-cortez-et-al-2009