

SOLVING DATA ANALYSIS CHALLENGES WITH A VERSATILE ML PLATFORM

Mini Machine Learning Platform Capstone Project

2024 April 16

Satinderjit Singh

Implementation: Technology Used

```
pyproject.toml
1 [tool.poetry]
2 name = "ml-framework"
3 version = "0.1.0"
4 description = "This project defines the framework for future ML projects"
5 authors = ["dlp <mossdet.detector@gmail.com>"]
6 readme = "README.md"
7
8 [tool.poetry.dependencies]
9 python = ">=3.10,<3.12"
10 jupyterlab = "4.1"
11 seaborn = "0.13"
12 scikit-learn = "1.4"
13 imbalanced-learn = "0.12"
14 optuna = "3.5"
15 liveplotloss = "0.5.5"
16 dash = "2.15.0"
17
18
```

```
18
19 # Issue between poetry and tensorflow metadata since >=2.11
20 # This is a temporary workaround
21 # related to https://github.com/python-poetry/poetry/issues/8271
22 # Inspired from https://github.com/tensorflow/tensorflow/blob/adb39b04e9cb116df4659a7e2de9eea27e62f25c/tensorflow/tools/pip\_package/setup.py
23 tensorflow = {version = "2.15.0" }
24 tensorflow-macos = { version = "^2.13.0", platform = "darwin", markers = "platform_machine=='arm64'" }
25 tensorflow-intel = { version = "^2.13.0", platform = "win32" }
26 tensorflow-cpu = [
27     { version = "^2.13.0", platform = "linux", markers = "platform_machine!='arm64' and platform_machine!='aarch64'" },
28     { version = "^2.13.0", platform = "darwin", markers = "platform_machine!='arm64' and platform_machine!='aarch64'" },
29 ]
30 tensorflow-cpu-aws = { version = "^2.13.0", platform = "linux", markers = "platform_machine=='arm64' or platform_machine=='aarch64'" }
31 # https://github.com/tensorflow/tensorflow/blob/adb39b04e9cb116df4659a7e2de9eea27e62f25c/tensorflow/tools/pip\_package/setup.py
32 # https://github.com/python-poetry/poetry/issues/8271#issuecomment-1697740447
33 tensorflow-io-gcs-filesystem = [
34     { version = ">= 0.23.1", markers = "platform_machine!='arm64' or platform_system!='Darwin'" },
35     { version = "< 0.32.0", markers = "platform_system == 'Windows'" }
36 ]
37
38
39 [tool.poetry.group.dev.dependencies]
40 black = "24.2.0"
41
42 [build-system]
43 requires = ["poetry-core"]
44 build-backend = "poetry.core.masonry.api"
45
```

Brief overview of the problem the project aims to solve:

Lack of a versatile machine learning platform for structured dataset analysis.

The platform's ability to handle structured datasets of varying sizes and complexities can catalyze innovation and drive progress in data-driven fields.

Importance of developing a solution to address this problem.

By providing a centralized tool for data manipulation, preprocessing, visualization, classification, regression modeling, and clustering, the platform aims to enhance the efficiency and accuracy of data analysis workflows.

Challenges Faced and Solutions

Challenges

Lack of knowledge in various technologies like Docker, Logging, unit-testing

Integrating various libraries and other tools

Ensuring Scalability and Flexibility

Implementing complex machine learning algorithms

Solutions

Overcame these by asking Omar questions, watching youtube videos

Again asking Omar questions and also thanks to Daniel for the poetry file

Adopted a Object-Oriented design to overcome scalability and adaptability

overcame thorough research and experimentation

Marketing/Selling the Product

Utilize digital advertising platforms, such as Google Ads and social media advertising

Create informative blog posts, tutorials, and case studies highlighting the capabilities and benefits of the mini ML platform.

partnerships with educational institutions, research organizations, and industry associations to promote the mini ML platform

Host live demo sessions and webinars to showcase the platform's features and functionalities, allowing potential customers to interact with the product and ask questions.

Unique Selling Points & Value Proposition:

Versatility: Highlight the platform's versatility in handling diverse structured datasets and performing various data analysis tasks, including preprocessing, visualization, classification, regression modeling, and clustering.

User-Friendly Interface: Emphasize the user-friendly interface of the mini ML platform, designed to streamline complex data analysis workflows and provide an intuitive user experience for data analysts of all skill levels.

Scalability & Performance: Showcase the platform's scalability and performance capabilities, enabling efficient processing of large datasets and delivering accurate analysis results in a timely manner.

Target Audience

Data analysts, Data scientist

,Researchers

Small businesses seeking efficient solutions for structured dataset analysis.

Project Demo

Q&A