pandas==2.1.4

numpy==1.26.2

scikit-learn==1.3.2

streamlit==1.29.0

nltk==3.8.1

joblib==1.3.2

plotly==5.18.0

```

<antArtifact identifier="README-md" type="text/markdown" title="Project README">

# MovieLens Prediction Project

## Project Overview

A machine learning project for movie recommendations and rating predictions using the MovieLens dataset.

## Setup Instructions

1. Clone the repository

2. Create a virtual environment

```

python -m venv venv

venv\Scripts\activate # On Windows

```

3. Install dependencies

```

pip install -r requirements.txt

```

4. Download NLTK Resources

```python

import nltk

nltk.download('punkt')

nltk.download('stopwords')

```

5. Prepare Data

- Place MovieLens CSV files in the `data/` directory

6. Train Models

```

python -m src.model\_training

```

7. Run Streamlit App

```

streamlit run app.py

```

## Features

- Movie Rating Prediction

- Genre Analysis

- Similar Movie Recommendations

## Technologies

- Python

- Pandas

- Scikit-learn

- Streamlit

- NLTK

## Contributors

[Your Name]

```

<antArtifact identifier="data-folder-readme" type="text/markdown" title="Data Folder README">

# MovieLens Dataset

## Required Files

1. movies.csv

2. ratings.csv

3. links.csv

4. tags.csv

5. genome-scores.csv

6. genome-tags.csv

## Download Instructions

Download these files from the MovieLens dataset:

- https://grouplens.org/datasets/movielens/latest/

```

Detailed Steps for Complete Setup:

1. Activate Virtual Environment

```bash

cd D:\movies\_prediction\project

python -m venv venv

venv\Scripts\activate

```

2. Install Dependencies

```bash

pip install -r requirements.txt

```

3. Download MovieLens Dataset

- Go to https://grouplens.org/datasets/movielens/latest/

- Download `ml-latest.zip`

- Extract CSV files to `D:\movies\_prediction\project\data`

4. Verify Dataset

- Check that `data` folder contains:

- movies.csv

- ratings.csv

- links.csv

- tags.csv

- genome-scores.csv

- genome-tags.csv

5. Train Models

```bash

python -m src.model\_training

```

6. Run Streamlit App

```bash

streamlit run app.py

```

### Potential Challenges and Solutions

1. \*\*NLTK Download Issues\*\*

If you face problems downloading NLTK resources:

```python

import nltk

nltk.download('punkt', download\_dir='./nltk\_data')

nltk.download('stopwords', download\_dir='./nltk\_data')

```

2. \*\*Memory Constraints\*\*

- For large datasets, consider sampling or using more memory-efficient techniques

- Add logging to track processing

3. \*\*Model Performance\*\*

- Experiment with different algorithms

- Use cross-validation

- Try hyperparameter tuning

### Best Practices

- Always use virtual environments

- Keep sensitive information out of version control

- Regularly update dependencies

- Add error handling in scripts

- Use logging for tracking processes

### Troubleshooting

1. Ensure all required libraries are installed

2. Check Python version compatibility

3. Verify dataset file paths

4. Use `print()` or logging for debugging