Gene Mutation Classification using Machine Learning

Project Overview

This project uses machine learning to classify genetic mutations (e.g., Missense, Nonsense, Silent) based on genomic features. The goal is to assist researchers and healthcare professionals in identifying mutation types efficiently.

Dataset

Source: DepMap Portal

Size: 1.2 million rows, 32 columns

Features: Chromosome, Variant_Type, Reference_Allele, Alternate_Allele, Start_Position, End_Position, etc.

Target: Variant_Classification (e.g., Missense_Mutation, Nonsense_Mutation)

Workflow

1. Data Preprocessing:

- Handled missing values using mean/median

- Encoded categorical features

- Scaled numerical values

2. Model Building:

- Used Random Forest Classifier from scikit-learn

- Hyperparameter tuning using GridSearchCV

3. Evaluation:

- Accuracy, Precision, Recall, F1-score

4. Deployment:

- Flask-based web app for prediction interface

Evaluation Metrics

Gene Mutation Classification using Machine Learning

Accuracy: 97.5%

Precision: 97.52%

Recall: 97.55%

F1 Score: 97.53%

Web App Features

- Input fields for mutation details

- Real-time classification of mutation type

- Simple and user-friendly interface

Technologies Used

- Python

- Scikit-learn

- Pandas, NumPy

- Flask

- HTML/CSS (basic)

- Pickle (for model saving)

How to Run Locally

1. Clone the repo

2. Install dependencies: pip install -r requirements.txt

3. Run the Flask app: python app.py

4. Open your browser and go to: http://127.0.0.1:5000/

Future Work

Gene Mutation Classification using Machine Learning

- Add deep learning models (e.g., CNNs for sequence data)

- Expand to disease prediction

- Add mobile-friendly UI

- Improve support for rare mutation types

References

- DepMap Portal

- scikit-learn documentation

- Flask documentation

- ChatGPT