Blood Transmutation Using Machine Learning Tools

IV CS Major Project Phase-I

A.Y. 2024-2025

- **Team Leader:** CHILUKURI SAI PRIYA - 21R21A6714
- **Team Members:**
  - BUCHIREDDY PALLI HARI KIRAN - 21R21A6712
  - GUMMA YASASVI - 21R21A6722
- **Guide:** MR. IRFAN BAGAWAN

Overview

- **Objective:** This project aims to safely alter blood types, improving transfusion safety and expanding the donor pool using machine learning.
- **Focus Areas:** We will study blood group antigens, genetic markers, and biochemical properties to accurately change blood types.
- **Reference:** Toward Universal Donor Blood: Enzymatic Conversion of A and B to O Type
  - **Authors:** Katsuhiko Kinoshita et al.
  - **Year:** 2020

Data Collection & Preprocessing

- **Blood Data:** Collect parameters like blood group, Rh factor, and hemoglobin levels to establish baseline characteristics.
- **Chemical Components Data:** Gather data on molecular formulas, concentrations, and reaction types to understand the chemical processes involved.
- **Conversion Methods Data:** Examine techniques, their effectiveness, and the chemicals required for blood conversion.
- **Experimental & Clinical Data:** Use lab results and clinical trials to validate and refine our models.

Machine Learning Models

- **Base Models:** Train Random Forest and XGBoost for initial predictions. Random Forest offers robustness, while XGBoost provides high performance.
- **Meta-Model:** A neural network will be trained on base model predictions to capture complex patterns and improve accuracy.
- **Ensemble Learning:** Combining multiple models enhances robustness and reduces overfitting, leading to better predictions.

Implementation Steps

1. **Train Base Models:** Start with Random Forest and XGBoost to generate predictions.
2. **Generate Predictions:** Combine predictions from base models to create a comprehensive dataset.
3. **Train Meta-Model:** Use a neural network to refine predictions from the combined dataset.
4. **Evaluate Performance:** Assess the final model's accuracy and reliability using test data.

Tools & Algorithms

- **Data Science Tools:** Pandas, NumPy, SciPy for data manipulation; scikit-learn and TensorFlow for machine learning.
- **Visualization:** Plotly and Seaborn for creating interactive and informative charts.
- **Deployment:** Use Django for web development, Docker for containerization, and Heroku for cloud deployment.

Data Formats & Sources

- **Structured Data:** CSV, Excel, and SQL databases for efficient data handling.
- **Unstructured Data:** Text from research papers and images for additional insights.
- **Sources:** Biomedical and chemical databases like PubChem for comprehensive data.

Evaluation & Conclusion

- **Model Robustness:** Ensemble learning improves prediction robustness and addresses model weaknesses.
- **Data Quality:** High-quality data is crucial for accurate predictions and reliable models.
- **Future Work:** Expand datasets, refine models, and explore new techniques to enhance prediction accuracy.

Thank You!

Thank you for your attention. We welcome any questions or feedback you may have.