Project Proposal: Longevity Prediction Tool

Group 2 – Team Rockstars

Project Overview:

People often struggle to connect lifestyle choices to their impact on lifespan. This project aims to develop a Longevity Prediction Tool that helps individuals understand the effects of their lifestyle behaviors on their life expectancy. The tool will provide a personalized analysis of how various factors influence an individual's lifespan, going beyond traditional health assessment methods. Our hope is to motivate individuals to improve their health by allowing them to directly quantify possible positive lifestyle changes.

Current Solutions:

Individuals primarily rely on doctor visits, internet research, public awareness campaigns, and self-help resources to understand how their behaviors influence their health. Most of these solutions often provide limited, impersonal guidance, and lack transparency as to how they were derived. Doctor visits are infrequent, and the impact of certain lifestyle choices may not be adequately communicated.

Existing longevity prediction tools exist, such as the one available here, but they tend to only provide a static numerical output. Our longevity prediction tool will allow users to see detailed statistics describing how each of their lifestyle choices impacts their predicted lifespan.

Proposed Approach:

Our approach involves collecting a comprehensive set of features that could affect an individual's lifespan. We aim to use both macro-level features (nation, region, ethnicity, etc.) as well as personalized micro-level features (bad habits such as cigarettes, unhealthy diet, etc.) to give individuals a better understanding of the factors that may influence their lifespan. Through the integration of data from various sources, our aim is to develop a tool that provides individuals with valuable insights into how their lifestyle choices directly influence their life expectancy.

The tool we plan to develop will allow individuals to conduct a *lifespan audit*, offering a comprehensive overview of their health and lifestyle factors that influence longevity. Users can input their personal data and habits to receive insights on potential areas for improvement. Additionally, the tool features a map highlighting longevity hotspots, regions where people tend to live longer, along with insights gathered from these areas. Users can explore these insights to make informed choices for a healthier, longer life.

Measuring Success:

Success will be assessed by analyzing extensive research from multiple fields, establishing the expected range of impact that each feature should have on lifespan. We will develop and train a predictive regression model, evaluating its performance using an isolated test set. We will consider aggregated (e.g., national or city-level) and individual-level performance. Key success

metrics include the Mean Absolute Percentage Error (MAPE), which is suitable for regression tasks and easily understandable for non-technical audiences, and Mean Squared Error (MSE), which may be more valuable when training our model due to its ability to penalize large errors more than smaller ones.

Data Sources:
Our initial plan involves utilizing the following data sources:
1. Food Environment Atlas (API) - Data on food access indicators in the US
2. Data.gov (CSV/JSON) - Data on environmental factors such as disease and healthcare
3. Kaggle (CSV) - Labeled life expectancy data on the national level
4. ONS.gov (CSV) - Socioeconomic factors that affect lifespan
5. Healthdata.gov (CSV/JSON) - Lifespan by race ethnicity
6. Hofstede-insights (Web Scraping) - Cultural comparisons of countries
7. Meteomatics (API) - Weather data API to establish weather as a feature for longevity
8. Wikipedia (Web Scraping) - Lifespan data on the individual level

Timeline:
1. Backend Development: October 24 - November 7
   - Load and clean data to prepare for analysis by November 1
   - Isolate the features for our model and create training/testing datasets by November 3
   - Finalized data analysis and modeling code expected by November 5
   - Model testing and evaluation by November 7

2. Frontend Development: November 7 - November 14
   - Development of a web application with a React frontend and Flask backend

3. Presentation Preparation: November 14 - November 20

Team Member Roles:
Each team member will contribute to all aspects of the project but with the following specializations:
- Data Loading and Cleaning: Shuaiming + Mrinoy
- Modeling Process and Evaluation: Mrinoy + Nick
- Testing: Richmond Baker
- Frontend Development: Nick + Mrinoy
- Documentation: Richmond

This project will empower individuals with actionable insights, enabling them to make informed choices about their lifestyles and take control of their health, ultimately guiding them towards a longer, healthier life.