**Summer 2023 Project Draft**

**Cook’s County Treasurer’s Office – First Edition**

**June 7th, 2023**

**By Tinh Cao**

**General Information:**

*Summary:* The goal of this project is to develop a predictive model that considers of 6 years of property tax data ranging from 2017 to 2022 and provides insights for the office on issues such as early sign of late tax payment, sign of overpay and refund amount, sign of default in paying tax and need to put up into tax sale.

*Project Type:* New Initiative

*Project Start Date:* 05/15/2023

*Project End Date:* 08/06/2023

**Scopes:**

* **Objective 1: Predictive Modeling for Tax Sale:** Develop a predictive model to determine the likelihood of properties with specific characteristics being offered at the annual tax sale or the biennial scavenger tax sale. This model should consider relevant features such as property tax bill amounts, tax payments, assessed values, market values, and demographic details.
* **Objective 2: Revenue Forecasting:** Establish a machine learning-based revenue forecasting system to estimate expected tax revenues for properties in a specific municipality or township within 1 month, 3 months, and 6 months of the tax bill due date. This forecast should incorporate historical data and trends in payment timeliness.
* **Objective 3: Refund Estimation:** Construct a model to estimate potential tax refunds due to property tax assessment appeals or other causes that may lower the tax liability on properties after the taxes have been paid. The model should consider historical refund data, recent changes in tax laws, and other relevant factors.
* **Objective 4: Failure Mode Analysis:** Investigate the use of Machine Learning techniques to aid in Failure Mode and Effects Analysis within the Office. The goal is to detect and predict failure points in the property tax collection process, which could include inaccuracies in tax calculations, delays in bill processing, or errors in property assessments.
* **Objective 5: Data Visualization:** Develop a comprehensive data visualization that can display GIS mapping of properties, tax rates, demographics, and other relevant data. This would aid in the understanding and interpretation of patterns, correlations, and trends in property tax data.
* **Objective 6: Policy Decision Support:** Leverage the above models to inform policy decisions within the office and to provide estimations and analyses to local government units. The insights generated from these models should serve as robust decision support tools, informing and enhancing policymaking related to property tax.

**Assumptions:**

* Data is clean and well-structured.
* Features can be engineered toward objectives.
* Machine learning algorithms can leverage the dataset and make informative suggestions/insights.
* Open-source libraries do not violate any policy set up by the department.

**Risk and Mitigation Plans:**

* Structural risk: Successor teams may decide to use a different method/approach toward analysis.
* Scalability: Future projects may require distribution computation or enterprise infrastructure for a more complex/ large dataset.
* Misuse of predictive models: Potential of discriminatory outcomes.
* Data security: Risks of data breaches or unauthorized access.
* Data quality: Poor data, inconsistencies, inaccuracies could lead to poor model development.
* Dependency on predictive model used: Models are fallible, and relying on predictive model could lead to inaccuracies, or biased.
* Misinterpretation: Misinterpretation or misunderstanding of model could impair the decision-making process.
* Algorithm-biased: The choice of algorithm may lead to un unfair or discriminatory predictions.
* Regulatory Compliance: The implementation of the project could be subject to laws and regulations related to AI, fairness, and privacy.
* Technological change: Fast-paced AI and other machine learning algorithms development can make the project outdated.

**Success Criteria:**

1. Success Criteria:
   1. A minimum viable product (MVP) for Tax sale model:
      1. Data Preprocessing Tools: A set of tools to clean, preprocess, and integrate necessary data, including historical tax revenue data, payment timeliness, tax rates, and assessed values.
      2. Predictive Model: A machine learning model that leverages the preprocessed data to predict future tax revenues within specific time frames (e.g., 1 month, 3 months, 6 months).
      3. User Interface: A simple UI that allows users to select a municipality or township and view the predicted tax revenues for the selected time frames.
      4. Basic Reporting: A feature to generate simple reports summarizing the predicted revenues for the selected regions and time frames.
   2. MVP for Revenue Forecasting System model:
      1. Data Processing: Tools to handle and preprocess necessary data, including historical refund data, recent changes in tax laws, and other relevant factors.
      2. Predictive Model: An AI model that uses processed data to estimate potential tax refunds.
      3. User Interface: A straightforward interface that allows users to input specific parameters and get back an estimated refund amount.
      4. Simple Reporting: A functionality that allows the generated estimations to be exported for further analysis or record keeping.
   3. MVP for Basic Failure Mode Analysis model:
      1. Data Preprocessing: Tools for cleaning, preprocessing, and integrating necessary data related to the tax collection process.
      2. Failure Mode Analysis Model: An AI model that leverages the preprocessed data to identify potential failure points in the collection process.
      3. User Interface: A simple UI that presents the identified potential failures in a clear and understandable manner.
      4. Initial Alert System: A basic alerting mechanism that notifies relevant personnel when a potential failure point is detected.
2. Current State:
   1. Building foundation for future studies.
3. Target State:
   1. Project ease decision process with Failure Mode Analysis and Effect Analysis.
   2. Complete product for all models developed.

**Project Team:**

1. Team Coordinator: Tinh Cao
2. Team Members: Chris Chen, Shashank Parameswaran, Zainab Hasnain, Sohaib Syed

**Development Environment:**

The predictive model will be developed using Python on personal laptop, with the potential for expansion to enterprise-level infrastructure. The visualization of findings will be performed using Tableau.

**High Level Work Breakdown Structure**

1. Inception Phase: This phase involved defining the project scope, objectives, deliverables, stakeholders, risks, assumptions and constraints. The main deliverable of this phase was the project document that outlines the project vision, scope, objectives, approach, roles and responsibilities, communication plan and high-level schedule.
2. Elaboration Phase: This phase involves refining the project requirements, designing the predictive model architecture, selecting the appropriate data sources and tools, and preparing the data for analysis. The main deliverables of this phase are the requirements specification document, the model design document, the data dictionary document and the cleaned and transformed data sets.
3. Construction Phase: This phase involves developing, testing and validating the predictive model using Python, as well as creating visualizations and dashboards to present the findings. The main deliverables of this phase are the predictive model code, the test cases and results document, the validation report document and the Tableau dashboard files. The construction phase will be divided into three iterations of two weeks each, with progress check-ups at the end of each iteration.
4. Transition Phase: This phase involves delivering the project results to the company, conducting a project review and evaluation, and closing the project. The main deliverables of this phase are the project presentation, the project report, the project evaluation document and the project closure document.

**Milestones and Deliverables:**

1. Inception Phase (Weeks 1-3) – 05/15/2023 – 06/04/2023 - Completed.
2. Elaboration Phase (Week 4): 06/05/2023 - 06/11/2023 - Current

* Task 1: Identify and categorize necessary variables for predictive modeling
* Task 2: Import and clean property tax data using Pandas library
* Task 3: Conduct initial exploratory data analysis (EDA) with visualizations using Matplotlib/Seaborn
* Task 4: Define data pre-processing pipeline, including normalization/standardization, missing data imputation

1. Construction Phase (Weeks 5-10): 06/12/2023 - 06/18/2023
   1. Iteration 1 (Weeks 5-6): 06/12/2023 - 06/25/2023

* Task 5: Develop feature selection method to optimize model performance
* Task 6: Create initial predictive model using Scikit-learn
* Task 7: Define testing plan, sample size and validation procedure to test the functionality
* Task 8: Evaluate model performance using appropriate metrics (accuracy, precision, recall, etc.)
* Task 9: Refine model based on evaluation, adjust parameters or feature selection
* Task 10: Document findings, challenges and improvement points from Iteration 1
  1. Iteration 2 (Weeks 7-8): 06/26/2023 - 07/09/2023
* Task 11: Implement model refinements based on Iteration 1 feedback
* Task 12: Conduct secondary model testing using validation set
* Task 13: Evaluate refined model performance, note improvements from Iteration 1
* Task 14: Further refine model based on evaluation, adjust model complexity as needed
* Task 15: Document findings, challenges and improvement points from Iteration 2
* Task 16: Develop initial Tableau dashboard for visualization
  1. Iteration 3 (Weeks 9-10): 07/10/2023 - 07/23/2023
* Task 17: Implement final model refinements based on Iteration 2 feedback
* Task 18: Conduct final model testing using validation set
* Task 19: Evaluate finalized model performance, note improvements from Iteration 2
* Task 20: Test model with actual data, if available
* Task 21: Finalize and optimize Tableau dashboard based on model results
* Task 22: Document final model and prepare for transition

1. Transition Phase (Weeks 11-12): 07/25/2023 - 08/06/2023

* Task 23: Conduct training on model usage, interpretation and dashboard manipulation
* Task 24: Collect feedback, answer queries and make necessary final tweaks
* Task 25: Conduct a project review and evaluation to assess the project performance, outcomes, lessons learned and best practices
* Task 26: Review model with stakeholders, deliver project presentation using Tableau/Powerpoint.
* Task 27: Finalize all project documentation and wrap up project