**C964 Computer Science Capstone**

**Carolina Zerga**

**Western Governors University**

**A. Letter & Project Proposal**

Dear Seniors, Nontechnical Managers, and executives,

I am pleased to present this proposal for the implementation of a groundbreaking data product aimed at revolutionizing the decision-making process related to animal adoptions at the Austin Animal Center.

Summary of the Problem:

Many animals face uncertain futures upon entering shelters, and predicting their likelihood of adoption can significantly impact their outcomes. Our goal is to develop a robust data product that predicts, with high accuracy, whether a new animal will be adopted within six months of arrival.

Benefits to the Customer and Decision-Making Process:

This data product will empower decision-makers by providing actionable insights into adoption patterns and influencing factors. By leveraging predictive and prescriptive analyses, it will not only forecast adoption rates but also recommend strategies to enhance overall adoption success.

Outline of the Data Product:

Our data product, the Animal Shelter Adoption Prediction System, will utilize advanced predictive modeling techniques to assess adoption likelihood. It will incorporate traits, historical adoption patterns, and shelter duration to provide a comprehensive view of each animal's journey.

Data Used:

The primary data source will be the Austin Animal Center Intakes and Outcomes dataset, encompassing detailed information about animal traits, shelter stays, and adoption outcomes. This rich dataset will be pivotal in constructing accurate predictive models.

Objectives and Hypotheses:

Our objectives are to accurately predict animal adoption rates within six months and propose actionable strategies to improve adoption outcomes. We hypothesize that by understanding adoption patterns, we can implement targeted interventions for animals at risk.

Project Methodology:

We will commence with an in-depth data gathering phase, followed by rigorous predictive modeling. The methodology involves employing both predictive and prescriptive analyses to ensure a comprehensive understanding of adoption dynamics.

Funding Requirements:

To execute this project successfully, we estimate a budget of $50,000 to cover data acquisition, technology infrastructure, and personnel costs. Please Keep in mind that these figures are still an approximate, and actual costs may vary based on the specifics of the project and any unforeseen challenges.

Impact on Stakeholders:

The implementation of the Animal Shelter Adoption Prediction System will positively impact shelter operations, animal welfare, and the community. Shelters can proactively address adoption challenges, leading to more successful outcomes for animals and greater community engagement.

Ethical and Legal Considerations:

We are committed to handling sensitive data with the utmost care and adherence to ethical standards. All data will be anonymized and used strictly for the intended purpose. We will comply with all relevant legal requirements and ensure transparency in our practices.

Expertise Relevant to the Solution:

My extensive background in data analysis, predictive modeling, and project management uniquely positions me to lead this initiative successfully. With a proven track record in similar projects, I am confident in delivering actionable insights that drive positive change.

The Animal Shelter Adoption Prediction System holds great promise in transforming the lives of shelter animals and enhancing the effectiveness of adoption efforts. I welcome the opportunity to discuss this proposal further and address any questions or concerns.

Thank you for considering this proposal. I look forward to the possibility of bringing this impactful project to completion.

Sincerely,

Carolina Zega

**B. Executive Summary: Animal Shelter Adoption Prediction System**

Decision Support Problem or Opportunity:

The Animal Shelter Adoption Prediction System aims to address the challenge of predicting the likelihood of animal adoptions within six months of arrival at the shelter. By leveraging historical data, the system will guide decision-making processes, enabling shelters to optimize resource allocation and improve overall animal welfare.

Description of Customers and Fulfilling Their Needs:

Our primary customers are animal shelter managers and staff. This product fulfills their needs by providing an efficient decision support tool, empowering them to proactively manage shelter resources, enhance adoption strategies, and ensure better outcomes for animals under their care.

Existing Gaps in Data Products:

Current practices often lack predictive capabilities, hindering shelters from strategically managing resources. This data product addresses this gap by introducing predictive and prescriptive analytics, fostering a more proactive and informed approach to animal shelter operations.

Data Available or Needed:

We will gather comprehensive data from Austin Animal Center Intakes and Outcomes, including animal traits, intake/outcome patterns, and adoption histories. This rich dataset forms the foundation for predictive modeling and prescriptive analysis.

Methodology for Design and Development:

Our approach involves meticulous data collection, predictive modeling using Python, and prescriptive analysis to formulate strategies for improving adoption rates. The Lean Six Sigma methodology will guide project execution, emphasizing continuous improvement.

Deliverables:

The deliverables encompass a predictive model, prescriptive strategies, and a user-friendly interface for shelter staff. Documentation and training materials will accompany these deliverables.

Plan for Implementation:

Implementation will follow a phased approach, starting with data gathering and analysis, progressing to model development, and concluding with system deployment. Anticipated outcomes include improved adoption rates, streamlined operations, and enhanced animal welfare.

Validation and Verification Methods:

Validation will involve comparing model predictions with actual adoption outcomes. Stakeholder feedback and system testing will verify that the product meets requirements. Iterative testing and refinement will ensure continuous improvement.

Programming Environments and Costs:

Python will be the primary programming language, incurring minimal costs for libraries and tools. Human resources will include data scientists, analysts, and developers.

Projected Timeline:

Data Collection and Analysis: 2 months

Model Development: 2 months

System Deployment: 1 month

Testing and Refinement: Ongoing

Full Implementation: 6 months

Milestones, Start and End Dates, Resources:

Milestone 1: Data Collection (Start: 02/01/24, End: 04/01/2024, Resources: Data Analyst)

Milestone 2: Model Development (04/15/24, End: 06/15/24, Resources: Data Scientists)

Milestone 3: System Deployment (Start: 07/01/24, End: 08/01/24, Resources: Developers)

Milestone 4: Testing and Refinement (Start: 02/15/24, End: Ongoing, Resources: QA Team)

Milestone 5: Full Implementation (Start: 08/15/24, End: 02/15/25, Resources: Full Team)

This Animal Shelter Adoption Prediction System represents a transformative solution for shelters, combining cutting-edge technology with a commitment to animal welfare.

**D. Developed Product Documentation**

Business Requirements

The Animal Shelter Adoption Prediction System aims to enhance decision-making at Austin Animal Center by predicting animal adoptions within six months. This data product, catering to shelter managers and staff, seeks to optimize resource allocation and improve overall animal welfare. Through comprehensive data analysis, predictive modeling, and prescriptive analysis, the system aims to provide actionable insights. The user-friendly dashboard, equipped with three visualization types, ensures ease of interpretation. Non-functional requirements include scalability, reliability, performance optimization, and usability. Adherence to a budget of $50,000 and ethical/legal standards guides the project, assuming availability of the Austin Animal Center dataset. This summarizing vision underscores the commitment to efficient resource utilization, targeted adoption strategies, and positive impacts on animal welfare.

Raw and Cleaned Data

The primary dataset, titled "Data," contains raw information encompassing intake, outcome, combined, and final datasets utilized in the analytical process.

For transparency and traceability, the cleaned dataset is stored in the "Clean Data.csv" file within the project structure, alongside the associated code detailing the data cleaning procedures for reference and reproducibility.

Code Analysis

We'll use Python, pandas for data manipulation, scikit-learn for machine learning, and matplotlib/seaborn for visualization. (the codes are provided within the project)

A graph with blue squares

Description automatically generatedA graph of a chart

Description automatically generated with medium confidence

A graph of a dog breed

Description automatically generated

A graph of different types of cat breeds

Description automatically generated

A diagram of a network

Description automatically generated

Hypothesis Verification

Our focus on predicting animal adoptions within six months using historical shelter data, aligns with our initial hypothesis of creating an easily integrable and user-friendly data product. Through rigorous testing and analysis, our results consistently stayed within predicted bounds, validating our hypothesis. The visual data provided above illustrates the efficacy of our predictive and prescriptive methods. Our machine learning model demonstrated a 85% accuracy. The project's objective and purpose are in line with our predictions and initial plans, setting the foundation for ongoing improvements and advancements toward the ultimate goal of enhancing the animal adoption forecasting.

Effective Visualization and Reporting

We focused on Animal Shelter Adoption Prediction; we recognize the significance of effective storytelling through visualizations at every stage of our data-driven journey. Our project aims to predict whether a new animal entering the shelter will be adopted within six months, and visual elements play a pivotal role in shaping the narrative. (visuals are under “Code Analysis”)

Data Exploration and Preparation:

Our initial phase involves navigating the complexities of shelter data; which are employed to reveal patterns and traits in animal intakes, outcomes, and adoption history. These visualizations guide us through the data landscape, ensuring a thorough understanding as we lay the groundwork for successive analysis.

Data Analysis:

At the heart of our project, visualizations take center stage to communicate the story within the data. Detailed charts and intuitive infographics elucidate the details of adoption patterns and influential factors. Statistical visualizations help us highlight key findings, making the analysis accessible to stakeholders with varying technical backgrounds.

Data Summary and Adoption Likelihood Detection:

As we culminate our journey, visual storytelling becomes the bridge between raw data and actionable insights. Compelling dashboards succinctly encapsulate the predicted likelihood of adoption within six months, providing stakeholders with a clear visual roadmap. Our approach ensures that the final presentation not only informs but captivates, making the adoption phenomenon and its detection a vivid and memorable part of our narrative.

With meticulous design considerations and a dedicated focus on impactful narrative construction, our Animal Shelter Adoption Prediction initiative aims to enhance the comprehension and influence of our research outcomes. This effort nurtures a heightened appreciation for the insights derived from our data exploration and analysis processes. Notably, our predictive model has an accuracy rate of 86%. The interpretability is further accentuated through visual aids, such as the decision tree plot generated with sklearn, revealing gender, particularly whether the animal is an intact female, as the most influential feature in adoption prediction.

Accuracy Analysis

Three distinct machine learning models were employed: the k-nearest neighbor with an accuracy of 85.29%, random forest with an accuracy of 86.22%, and decision tree with 85.49%. The product's accuracy is formidable, reaching an impressive nearly 86%. This remarkable performance highlights the efficacy of our Animal Shelter Adoption Prediction model in delivering accurate predictions. The elevated accuracy instills confidence in the dependability of our product, affirming its capacity to furnish valuable insights into animal adoption outcomes.

Application Testing

In the development of our Animal Shelter Adoption Prediction project using Jupyter Notebooks, a top-down methodology was employed to systematically test the functionality of each module within the integrated system. This approach involved inspecting the input and output of each cell at the database level to verify the execution of expected results. To ensure accuracy, any modification made to a module necessitated the reset of the demonstration database, followed by the re-execution of the entire program. While this process proved more labor-intensive than traditional unit testing, it ensured the robustness and seamless integration of each module. The modular pipeline has undergone thorough validation, with each module consistently processing expected inputs and producing the anticipated outputs, affirming its intended behavior.

Application Files

Our Animal Shelter Adoption Prediction project shows detailed design, leveraging the flexibility of the Python programming language through the use of Jupyter Notebook. The entirety of our refined source code is encapsulated within a singular and comprehensively named file, "AnimalAdoptionPrediction." For seamless data management, the project employs a dedicated database file titled "AnimalAdoptionData.csv." This file is a repository of thoroughly curated data encompassing intake records, outcome details, aggregated datasets, and the final dataset utilized for our predictive modeling. This strategic organization not only reflects a commitment to professionalism but also ensures the strength and coherence of our predictive analytics application.

User Guide

The following section outlines the necessary procedures for installing and utilizing the program:

1. Install Python and Jupyter Notebook in your Windows computer
2. Open a command prompt to install the following: (please write the following)

* “pip install pandas”
* “pip install numpy”
* “pip install pickle”
* ”pip install matplotlib”
* “pip install scikit-learn”

1. Unzip the associated files and place them into a project folder
2. Open the file called “Animal Adoption Prediction”
3. Click on the “Run All” button on the top
4. Once it runs, you will be able to see all the coding, visuals, models and predictions.

**E. Sources**

* Animal intake raw data report:

[**https://data.austintexas.gov/Health-and-Community-Services/Austin-Animal-Center-Intakes/wter-evkm/about\_data**](https://data.austintexas.gov/Health-and-Community-Services/Austin-Animal-Center-Intakes/wter-evkm/about_data)

* Animal outcome raw data report:

**https://data.austintexas.gov/Health-and-Community-Services/Austin-Animal-Center-Outcomes/9t4d-g238/about\_data**