2023

BUSINESS PROPOSAL

CrowdDoing

PREPARED FOR:

Soumya Rajan

SMARTSIMULATE SOLUTIONS

PREPARED BY:

OLUWADAMILARE SHIJI, QIN REN, Fangji Chen

**Table of Contents**

**Introduction 2**

**Project Description 3**

Objectives 3

Requirements 3

Scope of Project and Out of Scope 3

Outcomes and Benefits 4

Deliverables 4

**Team Profiles 5**

**Execution 6**

Resources 7

Techniques 7

Tools 7

Execution Process 8

**Process Visualization 9**

**Potential Challenges 9**

Risk and Issues Management 9

Timeframe 9

Project Team 9

Model Selection 9

Communications 10

Scope Creep and Changes 10

**Code of Conduct 10**

Ethical Considerations 10

Information Management 11

**Milestones 12**

**Conclusion 12**

**INTRODUCTION**

* About the Client

CrowdDoing is a renowned global initiative that brings together a diverse community of volunteers, experts, and organizations to collaboratively address pressing social and environmental challenges. With a strong focus on collective impact, CrowdDoing aims to create sustainable solutions and drive positive change across various domains.

* Background of the Topic

Forests play a vital role in maintaining ecological balance and supporting biodiversity. However, various factors such as climate change, disease, and human activities have posed significant threats to the health and stability of forest ecosystems. Recognizing these challenges, our "CrowdDoing - Forest Health Improvement" project seeks to develop a method for accurately identifying forest regions with healthy and abnormal vegetation, providing valuable insights to support conservation efforts.

* Describe the Client’s need

As part of our collaboration with CrowdDoing, we understand the importance of enhancing their forest health monitoring capabilities. The organization requires an efficient and accurate means to analyze vast vegetation maps and distinguish between healthy and abnormal forest regions within provinces. By doing so, they aim to streamline their conservation efforts, allocate resources effectively, and make informed decisions for sustainable forest management.

* Project Rationale

The "CrowdDoing - Forest Health Improvement" project aligns perfectly with CrowdDoing's mission and objectives. Through the utilization of data science, machine learning, and data visualization techniques, we aim to provide them with actionable insights derived from the analysis of vegetation maps. This will not only enable a more comprehensive understanding of forest health but also help identify potential intervention areas to improve the overall ecological balance.

**PROJECT DESCRIPTION**

**Objective**

* Develop a Method for Identifying Forest Health: The primary goal is to devise an efficient method to distinguish between healthy vegetation (comprising live, fully developed trees, and bushes) and abnormal vegetation (including dead, dry trees, and debris) using vegetation maps for a selected Canadian province.
* Visualize and Analyze Vegetation Maps: The project aims to collect and visualize vegetation maps for the chosen Canadian province and gain insights into the visual appearance of healthy and abnormal vegetation on satellite images.
* Build a Classification Model: The project seeks to develop ground truth data and train a machine learning model to accurately classify forest regions as healthy or abnormal based on the vegetation map data.

**Requirements**

* Statistics: A solid understanding of statistical concepts is essential for analyzing and interpreting vegetation map data.
* Environmental and Forest Science: Knowledge of environmental and forest science principles will be crucial for comprehending forest health indicators and factors affecting vegetation.
* Data Science: Proficiency in data science methodologies, including data preprocessing, feature engineering, and model evaluation, is necessary for building an effective classification model.
* Machine Learning: Participants must be familiar with machine learning algorithms and techniques to develop a model capable of distinguishing between healthy and abnormal vegetation.
* Data Visualization & Report: Skills in data visualization using tools like Tableau will be valuable for presenting vegetation maps and analysis results effectively.
* Python and Tableau Knowledge: Competency in Python programming and Tableau software is necessary for data manipulation, analysis, and visualization tasks.

**Scope of Project and Out of Scope**

**In Scope**

* Collect and visualize vegetation maps for a selected Canadian province to identify healthy and abnormal vegetation regions.
* Develop ground truth data to train an efficient machine learning model for classifying healthy and abnormal forest regions.
* Train the machine learning model to classify forest regions based on the vegetation maps.
* Explore potential solutions to improve forest health using the results of the classification.
* Utilize data visualization techniques to present the findings and classification results effectively.
* Deliver Tableau vegetation maps, ground truth maps, source codes, and a detailed report with suggestions and implications.

**Out Scope**

* Detailed species identification within the forest regions.
* Real-time monitoring of forest health.
* Analysis and comparison of multiple Canadian provinces.
* Implementation of advanced machine learning techniques beyond the defined scope.
* Actual implementation of forest management solutions.
* Policy development related to forest health improvement.

**Outcomes and Benefits**

* Comprehensive analysis of forest regions based on vegetation maps, providing insights into healthy and abnormal areas.
* Increased understanding of the appearance of healthy and abnormal vegetation on satellite and aerial maps.
* Development of an efficient machine learning model to accurately classify forest regions as healthy or abnormal.
* Improved forest health management through the identification of potential solutions based on classification results.
* Enhanced data visualization techniques to effectively communicate the findings and classification outcomes.
* Valuable market research on the forestry industry, contributing to better decision-making and investment opportunities.
* Empowerment of environmental and forest science experts with data science and machine learning skills.
* Timely project management and delivery to meet project objectives and client requirements.
* Final report with visualized diagrams and implications for informed decision-making.

**Deliverables**

* Tableau vegetation maps showcasing healthy and abnormal forest regions.
* Ground truth maps for training and validating the machine learning model.
* Python program code and scripts used for data preprocessing and machine learning model training.
* A detailed report with suggestions and implications based on the classification outcomes.
* Visualized interactive dashboard presenting the project findings and classification results.

**TEAM PROFILES**

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Main Roles | Education | Experience/Skills |
| Qin Ren | Team lead & communicator | Bachelor’s degree in Economics | **Academic settings:** Knowledge in microeconomics, macroeconomics, stock market, business principles.  **Business settings:**  Audit Project Management: 5 years of experience as an Audit Project Manager, conducting financial statement audits, tax audits, market research, etc.  Financial Management: 3 years of experience as an accountant, specializing in financial management tasks.  **CPA Certification**: Certified Public Accountant (CPA) with expertise in accounting and finance. |
| Oluwadamilare  Shiji | Financial & Data Analyst | Bachelor’s degree in Metallurgical and Materials Engineering | 4years experience as a SCADA Database Engineer.  Skills: PLC Programming, Integration and Troubleshooting  SQL, Python, SSRS, Database Management. |
| Fangji Chen | Developer | Post Graduate Certificate in AIDI  Bachelor’s degree in EIE | **Experiences**:  Five years of hands-on experience as a Software Development Engineer in Test (SDET) specializing in mobile application testing and SDK development.  Six years of extensive experience in ETL (Extract, Transform, Load), data mining, and machine learning techniques focused on combating mobile advertisements antifraud.  **Skills**:  **Programming**: Strong programming skills in Python, JAVA, SQL, C#.NET, SHELL.  **Machine Learning Algorithms and Frameworks**: Proficient in various machine learning algorithms. Experienced in working with popular machine learning frameworks.  **Big Data Technologies**: Master of distributed computing frameworks and big data technologies. Experienced in processing and analyzing massive datasets with efficiency and scalability. |

**EXECUTION**

The initial phase of this project involves the project kickoff and clarifying the scope. During this stage, our team will work closely with our contact at CrowdDoing to schedule weekly meetings. These meetings will be used to discuss the project in detail, including specific requirements and expectations. We will also define the project's scope, both what is included (in scope) and what is not included (out of scope). Additionally, we will agree on a timeline for the project and the deliverables that will be provided.

Following the kickoff, the next step is data collection and preprocessing. We will be provided with all the necessary information and documents related to the project. In this stage, our focus will be on identifying and obtaining vegetation maps for the chosen Canadian province from reliable sources. The collected data will then undergo a thorough cleaning and preprocessing process to ensure it is suitable for analysis and model training. To gain a better understanding of the data, we will collaborate with environmental and forest science experts to identify relevant indicators and factors that affect vegetation health. Using Python notebooks, the data will be transformed and saved into a database or data warehouse using an SQL connection script. During this transformation, any outliers in the data will be detected and removed, and exploratory data analysis (EDA) will be conducted. We will also develop a method to create ground truth data to train and validate the machine learning model. The model itself will be implemented and trained using Python, with careful attention to avoiding overfitting and optimizing its performance. If needed, domain experts will manually label samples of vegetation regions as healthy or abnormal to support the model's training.

The final stage of the project involves visualization and reporting, which will be managed by the business analyst. The analyst will create an interactive dashboard using either Tableau or PowerBi to present the project's findings, including the classification results and other visualizations. Additionally, a comprehensive report will be generated summarizing the project's methodology, results, and implications. To ensure the project's quality and adherence to the defined scope, rigorous testing and validation will be performed on the dashboard, Python scripts, and all deliverables.

Upon completion of the project, we will deliver the raw data, finalized Tableau vegetation maps, ground truth maps, Python code, SQL scripts, and the detailed report to the client. We will also schedule a final meeting with the client to present the project's outcomes and address any questions or feedback they may have.

**Resources:**

* **Financial Data:**
* Financial Statements: Obtain financial statements from the investment company, including income statements, balance sheets, and cash flow statements. These can be acquired from the company's ERP system.
* Bank Reconciliation Statements: Obtain detailed bank reconciliation statements from different banks, including transaction records, deposits, withdrawals, and reconciled balances.
* Loan Contracts: Obtain copies of the loan contracts between the investment company and its borrowers, including terms, interest rates, repayment schedules, and any associated collateral.
* Lending Industry Data: Gather data from the lending industry through existing datasets from open government portals.
* **Software and Hardware:**
* Software: Python, Power BI, and Excel.
* These applications and programming languages will be used for data cleaning, analysis, and visualization.
* Hardware: Our personal computers will be used to run the software, store the data, and perform the analyses.

**Techniques:**

* Both the original data and cleaned data will be stored in a database using in-built SQL connection in python.
* This project will utilize the Python programming language to cleanse and analyze data.
* Python is extensively used in data analysis and will serve a critical role in the data cleaning and analysis stages in this project. With the help of Python, we will manipulate, clean, and analyze the financial data.
* We will employ various Python libraries, such as pandas for data manipulation, and NumPy for numerical computation, to manage large and complex data sets effectively.
* This is an appropriate Technique which will ensure that the data used in the analysis is accurate, consistent, and reliable, thereby enhancing the validity of the project outcomes.

**Tools:**

* This project will use Power BI, a business analytics tool, to create interactive visualizations and business intelligence reports.
* These visualizations will provide a clear and concise representation of the complex financial data analyzed during the project.
* By using the capability of Power BI, we can better understand the project findings and make informed decisions.
* The visualized dashboard will allow us to interact with the data and gain insights into specific areas of interest.

**EXECUTION PROCESS**

* Requirements Phase:

In the Requirements Phase, the project team will collaborate with the client, CrowdDoing, to understand their needs and define the project's objectives, scope, and deliverables. Weekly meetings will be scheduled to discuss the project in-depth, including any specific requirements and expectations. Key activities in this phase include:

- Defining the primary objective: Developing a method for identifying forest health and distinguishing between healthy and abnormal vegetation using vegetation maps for a selected Canadian province.

- Understanding data science, environmental science, and machine learning requirements.

- Gathering detailed information on vegetation maps and other relevant data sources.

- Identifying the project's scope and determining what is included (in scope) and what is excluded (out of scope).

- Defining the timeline and deliverables for the project.

* 2. Design Phase:

In the Design Phase, the project team will plan the overall approach and methodology for the project. This phase focuses on developing a clear strategy for data collection, preprocessing, model development, and visualization. Key activities in this phase include:

- Designing the data collection process and identifying reliable sources for vegetation maps.

- Planning the data preprocessing steps, including data cleaning, transformation, and outlier detection.

- Defining the ground truth data generation process for training and validating the machine learning model.

- Selecting appropriate machine learning algorithms for the classification task.

- Outlining the data visualization strategy using Tableau or PowerBi.

- Collaborating with environmental and forest science experts to gain insights into relevant indicators and factors affecting vegetation health.

* 3. Development Phase:

In the Development Phase, the project team will implement the planned approach and execute the various tasks. This phase involves collecting and preprocessing data, developing the machine learning model, and creating the interactive dashboard for visualization. Key activities in this phase include:

- Collecting vegetation maps for the selected Canadian province from identified reliable sources.

- Cleaning and preprocessing the data to make it suitable for analysis and model training.

- Collaborating with domain experts to manually label samples of vegetation regions for ground truth data.

- Implementing and training the machine learning model using Python, ensuring it avoids overfitting and is optimized for performance.

- Transforming and saving data into a database or data warehouse using SQL connection scripts.

- Developing the interactive dashboard using Tableau or PowerBi to present the project's findings, classification results, and visualizations.

* 4. Testing Phase:

In the Testing Phase, the project team will rigorously test and validate the various components of the project to ensure accuracy, reliability, and adherence to requirements. Key activities in this phase include:

- Conducting comprehensive testing of the interactive dashboard, Python scripts, and all deliverables.

- Validating the accuracy and performance of the machine learning model using appropriate metrics.

- Ensuring that the project meets the predefined scope and requirements outlined in the initial phase.

- Addressing and resolving any issues or bugs identified during testing.

* 5. Release Phase:

In the Release Phase, the project team will finalize and deliver all project outputs to the client, CrowdDoing. This phase involves presenting the project's outcomes and providing support for any further queries or feedback. Key activities in this phase include:

- Compiling and organizing the raw data, finalized Tableau vegetation maps, ground truth maps, Python code, SQL scripts, and the detailed report.

- Scheduling a final meeting with the client to present the project's outcomes, discuss the findings, and provide insights and recommendations.

- Delivering all project deliverables to the client and ensuring they have access to the interactive dashboard for further exploration.

- Offering post-project support for a specified period to address any technical issues or questions that may arise.

* By following this execution process, the "CrowdDoing - Forest Health Improvement" project will be successfully carried out, leading to valuable insights and contributions to conservation efforts and sustainable forest management.

**PROCESS VISUALIZATION**

|  |
| --- |
|  |

**POTENTIAL CHALLENGES**

**Risk and Issues Management**

When issues arise, my team will assess the situation and determine the optimal course of action. Should an issue impede our progress towards an objective, we will promptly notify the client, provide recommendations, and act according to their instructions. My team is dedicated to resolving issues, but this may affect our ability to meet anticipated timeframes.

**Timeframe**

This project is JIRA complaint and we will exercise our best efforts to respect deadlines and deliver the anticipated objectives.

**Project Team**

My team is confident in its ability to complete the project on time. However, we acknowledge that unforeseen circumstances such as exams, illness, or personal events may arise. In the event of any delay in the project's progress. The team leader will notify Ms. Natalia immediately and provide a plan outlining how we will meet the project's requirements.

**Model Selection**

Model selection is a crucial aspect of this Project. However, it can be time-consuming, and if a misfit is identified, it may not be possible to select a new model and complete the Process within the given timeframe. Therefore, each member of the Team will make every effort to ensure that the model selection is appropriate and efficient.

**Communications**

FRANDAN Investment Company will be fully committed to supporting the Project Team in achieving the objective. We understand the importance of this project and are willing to provide additional resources if required. My team is confident that together we can achieve the desired outcome.

**Scope Creep and Changes**

Scope creep is a common issue that arises when a project's scope is not adequately defined, documented, or controlled. It can also occur when new requirements are added to the project by either the client or the team. To mitigate this risk, our team will ensure that all requirements are clearly and accurately documented during the planning phase and that our resources are allocated accordingly. In the event of any scope changes, we will document the requirements, assess the impact on the project, notify the client, provide recommendations, and act according to their instructions. We are confident in our ability to manage scope creep and deliver a successful project.

**CODE OF CONDUCT**

**Ethical Considerations**

* We acknowledge our strict duty of confidentiality.
* We will ensure the observance of ethical practices demonstrating respect, honesty and dignity.
* We will thoroughly follow the Client's directives without deviation from the objectives.
* Value is always present in our thinking.
* Validated data will solely form the basis of our observations, conclusions, recommendations and decision-making.
* We undertake to use holistic analytics strategies and repeatable processes.
* We will thoroughly explain and document our analysis in detail for the Client. We will be transparent in all our activities.
* Moreover, we will check and recheck our research for quality, accuracy, completeness and integrity before presentation to maintain the validity and credibility of the results.
* We will proactively circumvent unethical behavior, such as exaggerating the results of our research.
* We will not falsely interpret, fabricate, embellish or otherwise misrepresent the data to validate our findings or change or omit details favoring making an analysis fit a hypothesis.
* We will be forthright and accountable if we make mistakes.
* We will tell the truth, even if it is bad news.
* Failure to observe deadlines is considered a breach of ethics.
* We have no conflict of interest and will not benefit from this Project.

**Information Management**

* We acknowledge that the data is an asset of any organization. The consequences and repercussions of unethical conduct when dealing with an organization's data can be significant and affect an organization's reputation, relationships and, ultimately, its revenues. Even the perception of unethical data handling has the power to undermine both internal and external trust.
* We give assurance that all the information provided is in our trust. It will remain private and protected from damage or alteration unless authorized.
* We will sign a Non-Disclosure Agreement that is acceptable to the Client and adhere to its conditions.
* We will notify the Client of the discovery of any sensitive information at any stage, namely personal identifying information or confidential information. We will make recommendations for addressing the issues and act according to their instructions.
* Original data and other information entrusted to us is stored in a secure location, such as SharePoint, and will remain unaltered.
* Copies of original data will be used for cleaning, discovery, manipulation and analysis.
* Naming conventions and version controls will segregate documentation, datasets and work product. Each will align with the applicable phase of the Project.
* All work will only be conducted through secure applications.
* All our workstations are protected by a password that is unknown to anyone other than the assigned user. No one has access to our workstations, including colleagues, family members and friends.
* All information sharing is secured from potential risks using encrypted channels, such as SharePoint.
* Discussions about the data, analysis and observations will only be amongst client-authorized collaborators and our course instructors.
* All information entrusted to us will only be used for its intended purposes unless specifically approved by the Client.

**MILESTONES**

|  |  |
| --- | --- |
| **Milestone** | **Complete By** |
| Acceptance of Project Charter | May 29, 2023 |
| Requirements Phase | Jun 11, 2023 |
| Design Phase | Jun 25, 2023 |
| Development Phase | July 8, 2023 |
| Testing Phase | July 15, 2023 |
| Release Phase | July 23, 2023 |
| Client Presentation | July 29, 2023 |
| Handover of the Deliverables Phase | Aug 3, 2023 |
| Project Close-Out Phase | Aug 11, 2023 |

**CONCLUSION**

In conclusion, we are confident in our ability to provide exceptional accounting services tailored to the specific needs of your investment company. Our qualified team, industry expertise, and commitment to excellence will ensure the successful completion of this project. We appreciate the opportunity to submit this proposal and look forward to the possibility of working with you to enhance your accounting processes and support your investment objectives.

Thank you for considering our services. If you have any questions or require further information, please do not hesitate to contact us. We value the opportunity to contribute to the success of your investment company.

Sincerely,

Qin Ren

Team Leader

SmartSimulate Solution