**Assignment 1 – A Data Science Project**

|  |  |
| --- | --- |
| **The Project:** | Predictive Maintenance for Industrial Equipment. |
| **Documentation and Issues in GitHub Repository a. README.md:** | The README.md file should describe the project, its goal and the technologies used:  The project aims to predict equipment failure before it occurs and reduce downtime and maintenance costs. Technologies such as Python, scikit-learn, and TensorFlow will be used to develop machine learning models that can analyze data from sensors on the equipment. |
| **Product Backlog:** | |
| **Project Goal:** | The project aims to develop a predictive maintenance system that can predict equipment failures before they occur, reducing downtime and maintenance costs. |
| **User Story 1:** | As a maintenance engineer, I want to collect data from sensors on the equipment so that I can use it to train the machine learning model. |
| **User Story 2:** | As a data scientist, I want to preprocess the collected data, so I can use it to train the machine learning model. |
| **User Story 3:** | As a data scientist, I want to develop a machine learning model to predict equipment failures and reduce downtime and maintenance costs. |
| **User Story 4:** | As a data scientist, I want to evaluate the performance of the machine learning model so that I can optimize it and achieve better results. |
| **User Story 5:** | As a software engineer, I want to implement the machine learning model in a web-based application so that maintenance engineers can use it to monitor equipment and predict failures. |
| **User Story 6:** | As a software engineer, I want to integrate the web-based application with the company's existing systems so that data from sensors can be collected and processed automatically. |
| **User Story 7:** | As a software engineer, I want to test the web-based application to ensure it works as expected. |
| **User Story 8:** | As a software engineer, I want to deploy the web-based application so that maintenance engineers can use it to monitor equipment and predict failures. |
| **Milestone:** | |
| **Sprint Goal:** | Develop and implement the machine learning model in a web-based application. |
| **Issues:** | User Stories 1 to 4 |
| **Description and Acceptance Criteria:** | |
| For each issue in the Milestone, the Description should explain what needs to be done, and the Acceptance Criteria should define when the issue is considered done. | |
| **User Story 1:** | As a maintenance engineer, I want to collect data from equipment sensors so I can use it to train the machine learning model. |
| **Description:** | Collect data from sensors on the equipment and save it in a file. |
| **Acceptance Criteria:** | * The data is collected from sensors on the equipment. * The data is held in a file. |
| **User Story 2:** | As a data scientist, I want to load the data into the project to start working on the analysis. |
| **Description:** | This task involves loading the data into the project by downloading it from a source or importing it from a file. The data should be loaded into a data structure suitable for analysis. |
| **Acceptance Criteria:** | * The data is loaded into the project. * The data is stored in a suitable data structure (e.g., pandas data frame). * The data is ready for analysis. |
| **User Story 3:** | As a data scientist, I want to clean the data so that I can have a better understanding of the data. |
| **Description:** | This task involves cleaning the data by removing any missing or invalid data, including removing rows or columns that contain missing or invalid data or imputing missing data. |
| **Acceptance Criteria:** | * Missing or invalid data is removed from the data set. * The cleaned data set is stored in a suitable data structure (e.g., pandas data frame). * The cleaned data set is ready for analysis. |
| **User Story 4:** | As a data scientist, I want to perform exploratory data analysis so that I can have a better understanding of the data. |
| **Description:** | This task involves performing exploratory data analysis to understand the data better, including creating visualizations, calculating summary statistics, and identifying trends or patterns in the data. |
| **Acceptance Criteria:** | * Summary statistics and visualizations are created. * The exploratory data analysis provides a better understanding of the data. * The exploratory data analysis results are documented. |
| **User Story 5:** | As a data scientist, I want to select a model for the analysis to start building a model. |
| **Description:** | This task involves selecting a model for the analysis; this may include reviewing the results of the exploratory data analysis and deciding which model is best suited for the data. |
| **Acceptance Criteria:** | * A model is selected for the analysis. * The rationale for choosing the model is documented. * The model is ready to be built. |
| **User Story 6:** | As a data scientist, I want to build a model to start making predictions |
| **Description:** | This task involves building the selected model; this may include preparing the data, fitting the model, and evaluating the model. |
| **Acceptance Criteria:** | * The model is built. * The model is evaluated, and its performance is documented. * The model is ready to make predictions. |
| **User Story 7:** | As a data scientist, I want to make predictions using the model to generate results. |
| **Description:** | This task involves making predictions using the built model; this may include preparing the data for prediction, making predictions, and evaluating the predictions. |
| **Acceptance Criteria:** | * Predictions are made using the model. * The predictions are evaluated, and the results are documented. * The results are ready to be presented. |
| **User Story 8:** | As a data scientist, I want to present the results to share the findings with others. |
| **Description:** | This task involves presenting the results to others, including creating visualizations, writing a report, or delivering a presentation |
| **Acceptance Criteria:** | * Results are presented clearly and concisely. * Results are accessible and understandable to others. * Results are ready to be shared with others. |
| **Working Agreement and Definition of Done:** | |
| **wa.md:** The Working Agreement should define how team members will work together and communicate with each other. It can include things like the communication channels that will be used, the working hours, and the expectations for responsiveness.  **dod.md:** The Definition of Done should define what constitutes a completed task, including code reviews, testing, and documentation. | |
| **Working Agreement:** | |
| **Communication:** | All team members will communicate clearly and effectively in person and online. If any team member is unavailable, they will share their absence and availability. |
| **Time Management:** | All team members will manage their time effectively and meet deadlines. If any team member cannot meet their deadline, they will communicate this as soon as possible. |
| **Collaboration:** | All team members will work together effectively and collaboratively. If any team member needs help, they will ask for it, and if they can help, they will offer it. |
| **Code Quality:** | All code will be well-documented and tested, following best practices. All team members will review the code before it is merged. |
| **Conflict Resolution:** | If any conflict arises, all team members will work together to resolve it promptly and constructively. |
| **Continuous Improvement:** | All team members will continuously look for ways to improve the project and their skills. |
| **Definition of Done (dod.md):** | * All tasks in the user story have been completed. * All acceptance criteria for the user story have been met. * The code has been reviewed and approved by at least one other team member. * The code has been tested and any bugs have been fixed. * The code has been merged into the main branch. * The user story has been updated with the completion date and any relevant information. |
| **Kanban Board:** | |
| The Kanban Board is a visual representation of the project's progress. The items in the Milestone should be placed in the "In Progress" column and moved to the "Done" column as they are completed. The Kanban board is a flexible tool that can be adapted to suit the needs of our project. The key is to keep it simple and focus on visualizing progress. | |
| **To build a Kanban board:** | |
| **Tool:** | We can use physical boards or digital tools such as Trello, Asana, or Jira to create our Kanban board. |
| **Define columns:** | Start by creating columns that match the stages of our project workflow, such as "To Do," "In Progress," and "Done." |
| **Add items:** | Put items, such as tasks or milestones, on sticky notes or cards and place them in the appropriate column. |
| **Set limits:** | Set limit on the number of items in each column to help maintain focus and prioritize work. |
| **Move items:** | As items are completed, move them from left to right through the columns until they reach the "Done" column. |
| **Review and improve:** | Regularly review the board and adjust to keep our workflow running smoothly. |
| **Pull Requests:** | |
| Pull Requests allow team members to merge their code changes into the project's main branch. When team members complete a task, they create a pull request to merge their code into the main branch. Other team members should review the pull request to ensure that the code is correct and meets the Definition of Done criteria before it is merged into the main branch. Overall, these steps are essential in organizing and tracking the progress of a data science project, ensuring that everyone is on the same page and that the project is completed efficiently and effectively. | |