1. **Project Overview**

Project name: Used car price prediction

Project goal: Develop a machine learning model that can accurately predict used car prices, and provide interactive visualization tools that allow users to gain insight into the basis for the model's predictions.

1. **Project phases and key tasks**

**2.1. Data collection and cleaning stage**

Task 1: Collect used car data set.

Task 2: Conduct preliminary data exploration and understanding.

Task 3: Handle missing values and abnormal data.

Task 4: Label encoding and feature engineering.

**2.2. Model development stage**

Task 5: Choose an appropriate machine learning algorithm.

Task 6: Build and train the model.

Task 7: Model evaluation and optimization.

**2.3. Visualization tool development stage**

Task 8: Design a prototype of an interactive visualization dashboard.

Task 9: Implement the front-end and back-end functionality of the dashboard.

Task 10: Test the interactivity and performance of the visualization tool.

**2.4. Final testing and optimization phase**

Task 11: Overall system testing.

Task 12: Conduct testing and feedback.

Task 13: Solve potential problems and make final optimizations.

1. **Key resources**

Data set: Collect a data set containing used car information (From kaggle.com)

Technical tools: Use Python, machine learning libraries (such as Scikit-Learn), visualization libraries (such as matplotlib, Tableau), web frameworks (such as Dash), etc.

1. **Timetable(Decide after communicating with the instructor)**

|  |  |  |
| --- | --- | --- |
| Phase/Task | Start Date | End Date |
| Data collection and cleaning |  |  |
| Model development |  |  |
| Visual tool development |  |  |
| Final testing and optimization |  |  |

1. **Milestone(The same as 4)**

Milestone 1: Data cleaning and feature engineering completed.

Milestone 2: Model training and evaluation completed.

Milestone 3: Visualization tool completed.

Milestone 4: Final testing and optimization completed.

1. **Ethics**

1. Be sure to obtain authorization when obtaining used car information.

2. Ensure that privacy regulations are followed when processing data sets and no data information is leaked.

3. Ensure that when training the model, feature selection and weight setting are fair and not biased towards any particular group.

4. Review model predictions to ensure they do not have an unfair or discriminatory impact on certain groups.

5. Improve the interpretability of the model, ensure that the decision-making process of the model can be explained, and reduce the opacity of the "black box" model.

6. Conduct regular ethical reviews of the project to ensure that the project complies with the latest ethical and regulatory requirements throughout the development cycle.

1. **Communication Plan**

Regular Meetings: Meet with mentors every week or two to share progress and solutions.

1. **Final Deliverable**

Deliverable 1: Machine learning model source code (GitHub repository)

Deliverable 2: Interactive dashboard for visualization tools

Deliverable 3: Final project report

1. **Background reading**
2. Gain a deeper understanding of the value and limitations of machine learning applications in the automotive industry, especially in the used car market and read relevant papers or articles to learn about past experiences and success stories from similar projects.
3. Investigate existing research on factors affecting second-hand car prices to understand the extent to which factors such as car age, mileage, brand, and model affect prices and heck out industry reports to learn about current market trends and key factors affecting used car prices.
4. Understand the advantages and disadvantages of different machine learning algorithms and choose the appropriate model based on project needs and research methods for hyperparameter tuning and performance evaluation of models to improve the prediction accuracy of the model.
5. Learn best practices for designing interactive visualization tools to ensure users can easily understand and use predictive models and read the literature on dashboard design and user experience to understand the best ways for users to interact with data.
6. Research relevant regulations and best practices for ethical issues and data privacy protection to ensure projects adhere to ethical and legal standards. Read research on data ethics and privacy protection to learn how sensitive information is handled in similar projects.
7. **Evaluation**
8. Data quality and cleaning effect evaluation:

Assess the quality of the data set, including completeness, accuracy, and consistency of the data. Analyze the effectiveness of the data cleaning process to ensure that missing values, outliers, and duplicate data are addressed.

1. Feature engineering and model training evaluation:

Evaluate the effectiveness of feature engineering to ensure that the selected features have a positive impact on the performance of the model. Perform cross-validation and model evaluation, using various metrics such as root mean square error, mean absolute error, etc. to evaluate the accuracy and generalization ability of the model.

1. Visual tool interactivity and user experience evaluation:

Conduct user testing to evaluate the ease of use and user experience of interactive visualization tools. Collect user feedback to understand user satisfaction with the tool and possible improvements.

1. Ethical Compliance Assessment:

Check whether the project complies with ethical standards and relevant regulations. Regular ethics reviews are conducted to ensure that projects adhere to ethical and legal guidelines throughout the development process.

1. Project Progress and Time Management Assessment:

Compare actual progress with planned progress to evaluate whether the project completes tasks at each stage on time. Analyze any potential delays or impediments and take appropriate corrective action.

1. **Software development methodology**

Incremental Development Model

1. **Reference**