**Internship Final Report**

**Student Name: Abirami.S  
University: Sastra Deemed University  
Major: Data Science  
Internship Duration: August 1st, 2023 - August 31st, 2023  
Company: ShadowFox  
Domain: Machine Learning  
Mentor: Mr. Hariharan  
Coordinator: Mr. Aakash**

**Objectives**  
During this internship, the primary objective was to build two machine learning models that dealt with multiple practical uses:  
  
**Image Tagging Model:** Construct a straightforward framework for classifying pictures into fundamental groups such as "cat," "dog," "car," and so on, using widely used frameworks like as TensorFlow and PyTorch.  
Create a model developed using machine learning to estimate the selling price of motor vehicles based on characteristics involving fuel type, years of service, showroom price, number of previous motorists, kilometer’s driven, seller type, and gearshift type.

**Task and Responsibilities**   
**Task 1:**

**Image Tagging Model Data Collection:** Collecting a collection of labelled pictures classified as "cat," "dog," "car," etc. Public datasets like the CIFAR-10 were utilized.   
**Data prior to treatment:** Enhanced model performance through the use of data prior to treatment tasks that include picture resizing, standardization, and data augmentation methods.   
**Model Selection and Training:** I chosen a Convolutional Neural Network (CNN) architecture and educated it with TensorFlow. To accelerate training and enhance accuracy, we utilized transfer learning to learn from a model that had been trained (ResNet).   
**Evaluation:** Model performance was determined using requirements like accuracy, precision, recall, and F1 score. I fine-tuned the model to enhance its performance.   
**Distribution:** Utilize TensorFlow Lite to prepare the model for the distribution as a mobile application, making sure real-world applicability.

**Task 2:**

**Car Selling Price Prediction Model Data Collection:** compiled historical data on automobile sales, including key variables such as fuel type, duration of service, showroom cost, number of previous proprietors, miles travelled, seller type, and gearbox type.   
**Data Preprocessing:** Managed missing values, encoded category variables, and scaled numerical properties to ensure data was prepared for modelling.   
**Feature Engineering:** Generated additional functions, such as the car's age, to enhance model performance.   
  
**Model Selection and Training:** Trained a number regression models (e.g., Linear Regression, Random Forest, Gradient Boosting Machines) and selected the best model according to cross-validation results.

**Evaluation:** Model accuracy was evaluated using assessment metrics that include Mean Absolute Error (MAE), Mean Squared Error, and R-squared. Examined feature importance in order to comprehend the primary drivers of automobile prices.   
**Deployment:** Developed a web application using Flask that enables consumers to provide car data and get a predicted selling price. The mathematical framework is being incorporated into this application to provide actual time forecasts.   
**Learning outcomes**   
I enjoyed direct experience utilising machine learning models to address real-world challenges.   
I realised the significance data pretreatment and feature engineering serve the purpose of improving model accuracy.   
Enhanced comprehension of TensorFlow and PyTorch for developing, training, and deploying machine learning models.   
Enhanced my comprehension of model evaluation metrics and their interpretation in the context of applications for business.   
Improving skills in solving problems by dealing with issues during the development and deployment of models.

**Challenges and Solutions.**  
**Data Imbalance in Image Tagging:** The dataset included an imbalance in the number of photos per category, that led to prejudiced predictions. Solution: To balance the information in the file, we employed oversampling techniques and data augmentation.   
**Overfitting in Price Prediction Model:** The original models had evidence that showed overfitting. Solution: To minimize overfitting, L1/L2 normalization and cross-validation was utilized.   
**Deployment Issues:** We enjoyed trouble in effectively implementing the models, especially for the mobile and internet-based apps. Solution: The models were optimized utilizing TensorFlow Lite and Flask, which leads to effortless integration and deployment.

**Conclusion:**  
This internship educated me much regarding how to build and deploy models based on machine learning in applications that are practical. The experience of working with real-world data, building models, and providing them in accessible formats has equipped me with essential skills for an eventual career in data science. The projects performed have practical applications, demonstrating the potential of machine learning throughout multiple industries.   
  
**Acknowledgements**   
I am grateful to my mentor, Mr. Hariharan, for his advice and support throughout the duration of my internship. I would additionally like to thank my coordinator, Mr. Aakash, for his support and encouragement. Ultimately, I am glad to the entire ShadowFox team for providing me this valuable an opportunity to acquire knowledge and improve.