**Used Car Price Prediction**

***Project report submitted in***

***partial fulfillment of requirement for the award of degree of***

**Department of Artificial Intelligence**

***by***

#### Mr.Shreyas Deogade Mr. Sarthak udapure

**Mr.Shiv Sharma Mr.Tanmay Tijare**

***Guide***

**Prof. Minakshi Ramteke**

Assistant Professor GHRCE, Nagpur



### Department of Artificial Intelligence

G H Raisoni College of Engineering, Nagpur

(An Autonomous Institute affiliated to Rashtrasant Tukadoji Maharaj Nagpur University, Nagpur)

Accredited by NAAC with “A+” Grade

Ranked 163rd by NIRF, in the Engineering Category for India Ranking 2022,

**April 2024**

**© G H Raisoni College of Engineering, Nagpur, Year 2024**

**DECLARATION**

We, hereby declare that the project report titled “Used Car Price Predication” submitted here in has been carried out by us towards partial fulfillment of requirement for the award of Department Of Artificial Intelligence. The work is original and has not been submitted earlier as a whole or in part for the award of any degree / diploma at this or any other Institution

/ University.

We also hereby assign to G H Raisoni College of Engineering, Nagpur all rights under copyright that may exist in and to the above work and any revised or expanded derivatives works based on the work as mentioned. Other work copied from references, manuals etc. are disclaimed.

|  |  |  |  |
| --- | --- | --- | --- |
| Name | Mobile\_No | Email\_ID | Signature |
| Shreyas Deogade | 7499710788 | Shreyas.deogade.ai@ghrce.raisoni.net |  |
| Sarthak Udapure | 7709316844 | [Sarthak.udapure.ai@ghrce.raisoni.net](mailto:Sarthak.udapure.ai@ghrce.raisoni.net) |  |
| Shiv Sharma | 8788557729 | Shiv.sharma.ai@ghrce.raisoni.net |  |
| Tanmay Tijare | 9730933037 | tanmay.tijare.ai@ghrce.raisoni.net |  |

**Place: Nagpur Date**

## CERTIFICATE

The project report entitled as “**Used Car Price Prediction”** submitted by **Shreyas Deogade, Sarthak Udapure, Shiv Sharma, Tanmay Tijare** for the award of Department of Artificial Intelligence has been carried out under my supervision. The work is comprehensive, complete and fit for evaluation.



#### Prof. Minakshi Ramteke

#### Guide

**Assistant Professor**

Department of Artificial Intelligence G H R C E, Nagpur

#### Dr.Mangala Madankar Head of Department

Department of Artificial Intelligence G H R C E, Nagpur

#### ` Dr. Sachin Untawale Director

GHRCE, Nagpur

## ACKNOWLEDGEMENT

I would like to express my profound gratitude to Mrs. Mangala Madankar Ma'am, Head of Artificial Intelligence department, and Mr. Dr. Sachin Untawale Sir (Dean) of G. H. Raisoni College of Engineering (Autonomous College) in Rashtrasant Tukdoji Maharaj University (RTMNU) for their contributions to the completion of my project titled "Used Car Price Prediction".

I attribute the success to my project guide Mrs. Minakshi Ramteke whose tireless efforts, timely and esteemed guidance and encouragement enabled me to complete this challenging job within stipulated time period.

I am deeply indebted to the director of our esteemed institution Head of the department Prof. Mangala Madankar for the facilities provided to me. Their ever encouraging attitude, constant inspiration and whole hearted cooperation will always be remembered.

Finally I acknowledge the teaching and non teaching staff of the Department of Artificial Intelligence for direct or indirect help given to me for completing this project and for providing consistent encouragement.

I would like to acknowledge that this project was completed entirely by Shreyas Deogade, Sarthak Udapure, Shiv Sharma, Tanmay Tijare and not by someone else. With a deep sense of gratitude.

Projectee Projectee

Signature Signature

Shreyas Deogade Sarthak Udapure

Projectee Projectee

Signature Signature

Shiv Sharma Tanmay Tijare

**INDEX**

|  |  |  |
| --- | --- | --- |
| **Sr.No** | **Topics** | **Page No** |
| **1** | **Introduction** | **8** |
| **2** | **Objectives** | **11** |
| **3** | **Literature survey** | **12** |
| **4** | **PROPOSED METHODOLOGY/ SYSTEM ARCHITECTURE** | **15** |
| **5** | **Hardware / software specification** | **16** |
| **6** | **Implementation** | **17** |
| **7** | **Conclusion** | **24** |
| **8** | **References** | **25** |

**ARCHITECTURE INDEX**

|  |  |  |
| --- | --- | --- |
| Sr.No | Topic | Page.No |
| **1** | Proposed methodology/ system architecture | 7 |
| **2** | Flow of the project | 8 |

**ABSTRACT**

Title: Predicting Used Car Prices: A Machine Learning Approach

Abstract:

The used car market is a complex ecosystem influenced by various factors such as vehicle age, mileage, brand reputation, and market demand. Predicting accurate prices for used cars is crucial for both buyers and sellers to make informed decisions. This paper proposes a machine learning approach for predicting used car prices based on relevant features.

The dataset used in this study comprises various attributes including make, model, year of manufacture, mileage, fuel type, transmission type, and location. Feature engineering techniques are employed to preprocess and extract meaningful information from the dataset.

Several machine learning algorithms, including linear regression, decision trees, random forests, and gradient boosting, are implemented and compared for their performance in predicting used car prices. The models are trained and evaluated using techniques such as cross-validation and mean absolute error to ensure robustness and accuracy.

Experimental results demonstrate the effectiveness of the proposed approach in predicting used car prices with high precision. The model's ability to handle diverse features and provide reliable price estimations contributes to enhancing transparency and efficiency in the used car market. This research provides valuable insights for stakeholders involved in buying, selling, and valuing used cars, facilitating better decision-making and improving overall market dynamics.

**Keywords**: 1. Used car market

2. Price prediction

3. Machine learning

4. Feature engineering

5. Linear regression

6. Decision trees

7. Random forests

8. Gradient boosting

9. Feature analysis

## 1. INTRODUCTION

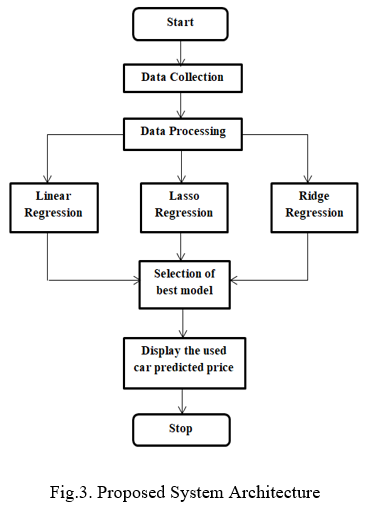
The used car market represents a significant sector of the automotive industry, characterized by its dynamic nature and diverse influencing factors. Unlike new car sales, where prices are relatively standardized and transparent, pricing used cars poses a greater challenge due to the wide range of variables involved. Factors such as vehicle age, mileage, brand reputation, market demand, and regional variations all contribute to the complexity of determining a fair and accurate price for a used car.

In recent years, the advent of machine learning techniques has provided new avenues for addressing this challenge. By leveraging vast datasets and sophisticated algorithms, it becomes possible to analyze historical trends and patterns, thereby enabling more precise predictions of used car prices. This not only benefits buyers and sellers seeking to make informed decisions but also enhances overall market efficiency and transparency.

This paper aims to explore the application of machine learning in predicting used car prices. We will delve into the methodologies involved, including feature engineering, model selection, and evaluation techniques. By examining the performance of various machine learning algorithms such as linear regression, decision trees, random forests, and gradient boosting, we seek to identify the most effective approach for accurately estimating used car prices.

Furthermore, we will discuss the implications of our findings for stakeholders in the used car market, including consumers, dealerships, and financial institutions. By providing insights into the factors driving used car prices and the reliability of predictive models, our research aims to empower stakeholders to make more informed decisions and navigate the complexities of the used car market more effectively. Ultimately, our goal is to contribute to the ongoing evolution of the automotive industry by leveraging the power of data and machine learning to enhance market efficiency and transparency in the realm of used car sales.

# ARCHITECTURE



## OBJECTIVE

The primary objective of this project is to develop a machine learning-based system for accurately predicting used car prices. This involves:

1. Improving Market Transparency: By providing reliable price estimations, the system aims to enhance transparency in the used car market, empowering buyers and sellers with the information they need to make informed decisions.

2. Enhancing Decision-Making: The system seeks to assist stakeholders, including consumers, dealerships, and financial institutions, in making more informed decisions related to buying, selling, and valuing used cars.

3. Optimizing Market Efficiency: By leveraging machine learning techniques to predict prices based on historical data and relevant features, the system aims to streamline transactions and improve overall market efficiency.

4. Mitigating Information Asymmetry: By leveraging data-driven insights, the system aims to mitigate information asymmetry between buyers and sellers, reducing uncertainty and facilitating fairer transactions.

5. Facilitating Pricing Strategies: The system aims to provide valuable insights into the factors driving used car prices, enabling stakeholders to develop more effective pricing strategies and adapt to market dynamics.

## 3.LITERATURE SURVEY/STUDY OF EXISTING SOLUTIONS/PRODUCTS

Literature Survey/Study of Existing Solutions/Products:

In recent years, there has been a surge in the development of solutions and products aimed at predicting used car prices using machine learning techniques. These solutions cater to various stakeholders in the automotive industry, including consumers, dealerships, and financial institutions. Here, we provide an overview of some prominent existing solutions and products:

1. Kelley Blue Book (KBB):

- Kelley Blue Book is a well-established platform that provides pricing information for new and used cars. They offer a range of tools and resources, including a used car valuation tool based on factors such as make, model, year, mileage, and condition. While KBB's pricing estimates are widely used and trusted, the underlying methodology for price prediction may not be explicitly disclosed.

2. TrueCar:

- TrueCar is another popular platform that offers pricing information and car-buying services. They leverage data analytics and machine learning algorithms to provide users with personalized price estimates for both new and used cars. TrueCar's pricing predictions are based on factors such as historical sales data, market trends, and individual vehicle attributes.

3.CarGurus:

- CarGurus is an online automotive marketplace that offers pricing insights and research tools for buyers and sellers. Their Price Analysis tool provides users with market-based pricing estimates for used cars, taking into account factors such as mileage, location, and vehicle history. CarGurus employs machine learning algorithms to analyze large datasets and generate accurate price predictions.

4. Autotrader:

- Autotrader is a leading online marketplace for buying and selling new and used cars. They offer a range of pricing tools and resources, including a Price Checker tool that provides users with estimated values for used cars based on factors such as make, model, year, mileage, and trim level. Autotrader's pricing estimates are powered by proprietary algorithms and market data analysis.

5. Edmunds:

- Edmunds is a trusted resource for automotive information and pricing. They provide users with pricing insights and research tools, including a True Market Value (TMV) pricing tool for used cars. Edmunds' TMV estimates are based on factors such as market trends, dealer inventory, and historical sales data, utilizing machine learning techniques to generate accurate price predictions.

6. Local Dealerships and Financial Institutions:

- Many local dealerships and financial institutions offer their own pricing tools and resources for estimating used car prices. These solutions may leverage machine learning algorithms and data analytics to provide personalized pricing estimates based on factors such as market demand, inventory levels, and financing options.

4. **PROPOSED METHODOLOGY/ SYSTEM ARCHITECTURE**

1. Data Acquisition and Preprocessing:

- Gather a comprehensive dataset containing information on used cars from various sources such as online marketplaces, dealership records, and public repositories.

- Preprocess the data by handling missing values, encoding categorical variables, and scaling numerical features to ensure uniformity and consistency.

2. Feature Selection and Engineering:

- Conduct exploratory data analysis to identify relevant features that may influence used car prices, such as make, model, year, mileage, fuel type, and location.

- Engineer new features or transform existing ones to capture important relationships and patterns in the data, such as aggregating mileage by year or extracting brand reputation scores.

3. Model Development:

- Select appropriate machine learning algorithms for predicting used car prices, such as linear regression, decision trees, random forests, or gradient boosting.

- Split the dataset into training and testing sets to train the models on a subset of the data and evaluate their performance on unseen data.

- Tune the hyperparameters of the selected algorithms using techniques such as grid search or random search to optimize model performance.

4. Model Evaluation:

- Evaluate the trained models using appropriate evaluation metrics such as mean absolute error (MAE), mean squared error (MSE), or R-squared to assess their predictive accuracy.

- Perform cross-validation to validate the generalization ability of the models and mitigate overfitting.

5. Deployment and Integration:

- Deploy the trained model into a production environment where it can be accessed by users for real-time predictions.

- Integrate the prediction model with existing platforms or applications used by stakeholders in the used car market, such as online marketplaces or dealership management systems.

6. Monitoring and Maintenance:

- Monitor the performance of the deployed model over time and retrain it periodically with new data to ensure its continued accuracy and relevance.

- Address any issues or drifts in model performance through regular maintenance and updates to the methodology as needed.

## 

## 5. HARDWARE / SOFTWARE SPECIFICATION

Hardware/Software Specifications:

1. Hardware Requirements:

- Processor: Multi-core CPU (Intel Core i5 or higher recommended) to handle computational tasks efficiently.

- Memory (RAM): At least 8GB RAM to support data processing and model training tasks. Higher RAM capacity may be beneficial for handling large datasets.

- Storage: Sufficient disk space (at least 100GB) to store datasets, models, and related files.

- Graphics Processing Unit (GPU) (optional): A dedicated GPU with CUDA support (NVIDIA GeForce GTX 1060 or higher) can accelerate training and inference tasks for deep learning models.

2. Software Requirements:

- Operating System: Windows, macOS, or Linux-based operating systems are supported.

- Python: Python programming language (version 3.6 or higher) is required for implementing machine learning algorithms and data preprocessing tasks.

- Integrated Development Environment (IDE): Recommended IDEs include PyCharm, Jupyter Notebook, or VSCode for coding and experimentation.

- Libraries: Install essential Python libraries such as NumPy, pandas, scikit-learn, and TensorFlow or PyTorch for machine learning tasks. Additional libraries for data visualization (matplotlib, seaborn) and model evaluation (scikit-learn, TensorFlow, Keras) may also be utilized.

- Database Management System (DBMS) (optional): If dealing with large datasets, consider using a DBMS like MySQL or PostgreSQL for efficient data storage and retrieval.

- Deployment Platforms: Depending on the deployment strategy, familiarity with cloud computing platforms such as Amazon Web Services (AWS), Microsoft Azure, or Google Cloud Platform (GCP) may be beneficial for deploying and scaling the predictive model.

3. Internet Connectivity:

- Stable internet connectivity is required for accessing online data sources, cloud services, and software updates.

4.Version Control System (optional):

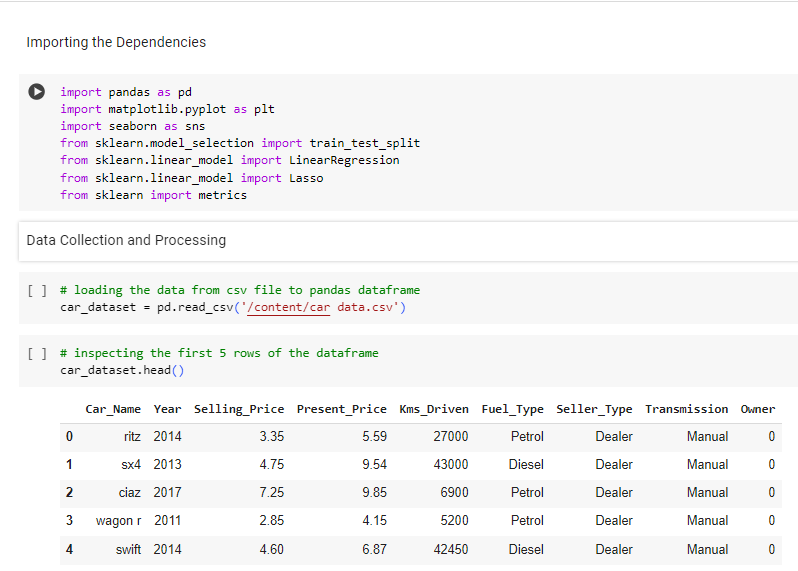
- Consider using a version control system such as Git and platforms like GitHub or GitLab for tracking changes, collaboration, and code management throughout the project lifecycle.

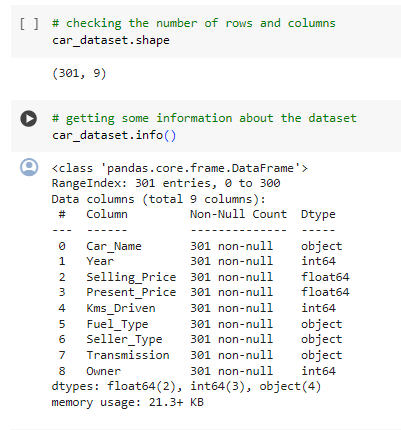
5. Documentation Tools:

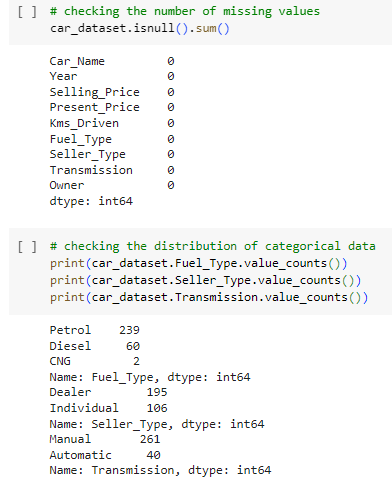
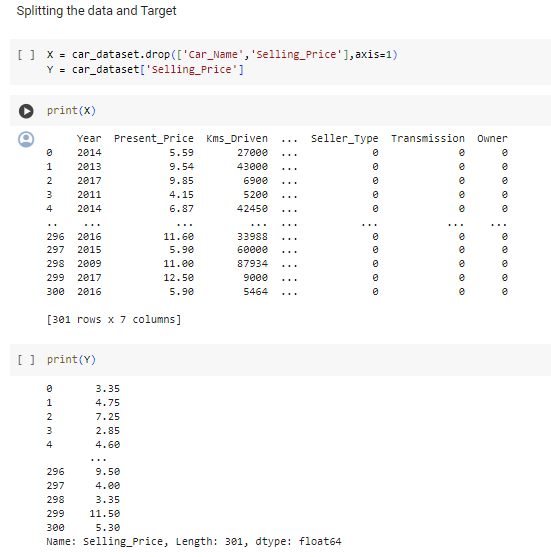
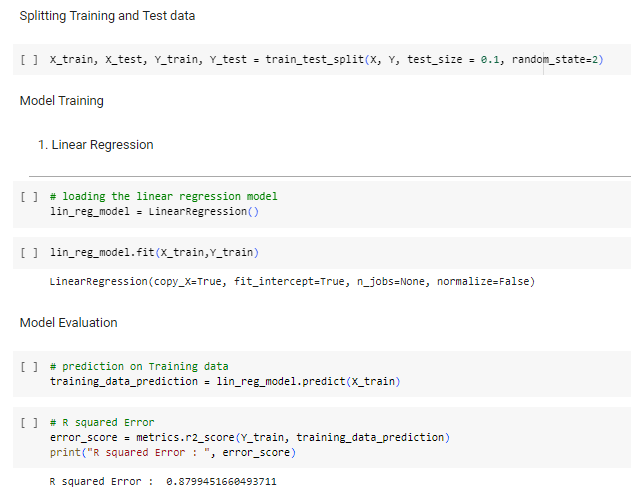
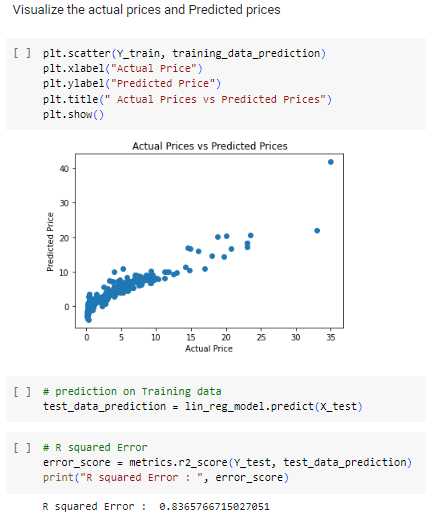
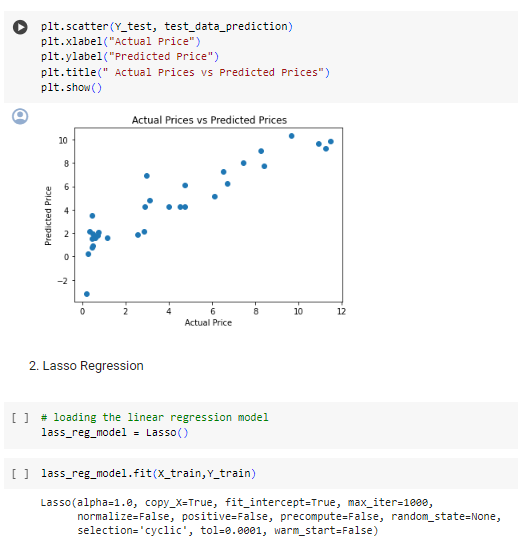
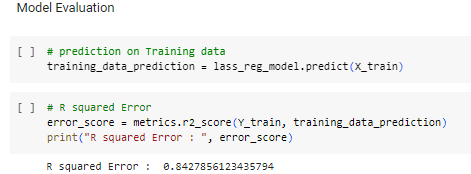
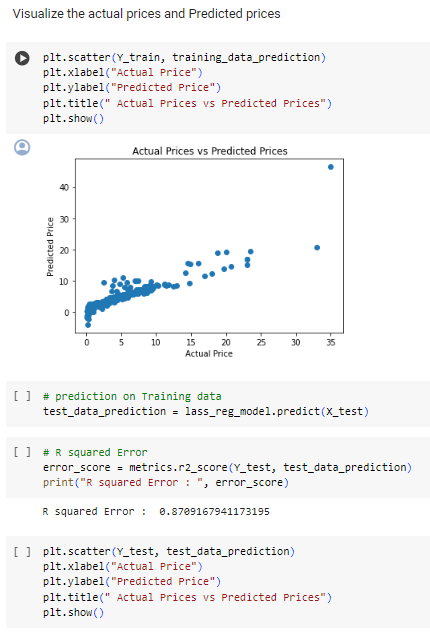
- Utilize documentation tools such as Jupyter Notebooks, Markdown, or LaTeX for documenting project progress, methodology, and findings.

By adhering to these hardware and software specifications, developers can effectively implement the proposed methodology for predicting used car prices using machine learning techniques. These specifications ensure adequate computational resources, software dependencies, and tools for efficient development, deployment, and maintenance of the predictive model.

**6. IMPLEMENTATION**

****



#### APP .PY

from flask import Flask,render\_template,request,redirect

from flask\_cors import CORS,cross\_origin

import pickle

import pandas as pd

import numpy as np

app=Flask(\_\_name\_\_)

cors=CORS(app)

model=pickle.load(open('LinearRegressionModel.pkl','rb'))

car=pd.read\_csv('Cleaned\_Car\_data.csv')

@app.route('/',methods=['GET','POST'])

def index():

    companies=sorted(car['company'].unique())

    car\_models=sorted(car['name'].unique())

    year=sorted(car['year'].unique(),reverse=True)

    fuel\_type=car['fuel\_type'].unique()

    companies.insert(0,'Select Company')

    return render\_template('index.html',companies=companies, car\_models=car\_models, years=year,fuel\_types=fuel\_type)

@app.route('/predict',methods=['POST'])

@cross\_origin()

def predict():

    company=request.form.get('company')

    car\_model=request.form.get('car\_models')

    year=request.form.get('year')

    fuel\_type=request.form.get('fuel\_type')

    driven=request.form.get('kilo\_driven')

    prediction=model.predict(pd.DataFrame(columns=['name', 'company', 'year', 'kms\_driven', 'fuel\_type'],

                              data=np.array([car\_model,company,year,driven,fuel\_type]).reshape(1, 5)))

    print(prediction)

    return str(np.round(prediction[0],2))

if \_\_name\_\_=='\_\_main\_\_':

    app.run()

### 

## CONCLUSION

In summary, developing a machine learning system for predicting used car prices offers significant benefits for stakeholders. By leveraging data analytics and advanced algorithms, decision-making in the automotive industry can be enhanced, market transparency improved, and information asymmetry reduced. Adequate hardware and software specifications are crucial for successful implementation. Overall, this approach holds promise for creating a more efficient and equitable used car marketplace.

## REFERENCES

[1] S. Pudaruth, “Predicting the Price of Used Cars using Machine Learn-

ing Techniques,” International Journal of Information & Computation

Technology, vol. 4, no. 7, pp. 753–764, 2014.

[2] N. Kanwal and J. Sadaqat, “Vehicle Price Prediction System using Machine Learning Techniques,” International Jounal of Computer Applications, vol. 167, no. 9, pp. 27–31, 2017.

[3] S. Peerun, N. H. Chummun, and S. Pudaruth, “Predicting the Price of

Second-hand Cars using Artificial Neural Networks,” The Second

International Conference on Data Mining, Internet Computing, and

Big Data, no. August, pp. 17–21, 2015.

[4] N.Sun, H. Bai, Y. Geng, and H. Shi, “Price evaluation model in

second-hand car system based on BP neural network theory,” in 2017

18th IEEE/ACIS International Conference on Software Engineering,

Artificial Intelligence, Networking and Parallel/Distributed

Computing (SNPD), jun 2017, pp. 431–436.

[5] G.Rossum, “Python Reference Manual,” Amsterdam, The

Netherlands, The Netherlands, Tech. Rep., 1995.

[6] A. K. Elmagarmid, P. G. Ipeirotis, and V. S. Verykios, “Duplicate

Record Detection: A Survey,” IEEE Transactions on Knowledge and

Data Engineering, vol. 19, no. 1, pp. 1–16, jan 2007.

[7] G.Chandrashekar and F. Sahin, “A survey on featureselection

methods,” Computers & Electrical Engineering, vol. 40, no. 1, pp.

16–28, 2014. [Online]. Available:

http://www.sciencedirect.com/science/article/pii/S0045790613003066

[8] M.C.Newman,“Regression analysis of log-transformed data:

Statistical bias and its correction,” Environmental Toxicology and

Chemistry, vol. 12, no. 6, pp. 1129–1133, 1993. [Online]. Available:

http://dx.doi.org/10.1002/etc.5620120618