Journal of **N**onlinear **A**nalysis and **O**ptimization
Vol. 15, Issue. 1, No.6 : 2024
ISSN : **1906-9685**

OPTIMIZING UBER: A COMPARATIVE STUDY OF ADVANCED REGRESSION MODELS FOR PRICE PREDICTION

Suman Baghel Department of Computer Science and Technology Shri Shankaracharya Institute of Professional Management and Technology
Yukti Jain Department of Computer Science and Technology Shri Shankaracharya Institute of Professional Management and Technology

Abstract
This study investigates five distinct regression models, namely Multiple Linear Regression, Ridge Regression, Lasso Regression, Elastic-Net Regression, and Polynomial Regression, the latter utilizing a polynomial order of 5. Our research outcomes underscore the supremacy of Polynomial Regression, particularly when implemented with a fifth-order polynomial, as it consistently outperforms its counterparts. Nevertheless, it is noteworthy that conventional multiple regression algorithms, encompassing Multiple Linear Regression, Ridge Regression, and Lasso Regression, offer comparable performance metrics.
Keywords: Multiple Linear Regression, Ridge Regression, Lasso Regression, Elastic-Net Regression, Polynomial Regression.

1. Introduction

1.1 Multiple linear regression

Multiple linear regression is a statistical technique employed to establish relationships between a dependent variable (Y) and multiple independent variables (X1, X2, X3, ... Xn) using the following expression:
$Y=\beta_0+\beta_1*X_1+\beta_2*X_2+...+\beta_n*X_n+\varepsilon$ [7]

1.2 Ridge regression

Ridge Regression is an extension of linear regression that introduces L2 regularization to mitigate issues like multicollinearity and overfitting. The Ridge Regression formula is as follows:
$Y=\beta_0+\beta_1*X_1+\beta_2*X_2+...+\beta_n*X_n+\varepsilon+\lambda*(\beta_1^2+\beta_2^2+...+\beta_n^2)$ [6]

1.3 Lasso regression

Lasso Regression is a regression technique that introduces L1 regularization to the linear regression model. Lasso Regression modifies the linear regression formula by adding an L1 regularization term (λ) to the least squares objective function:
$Y=\beta_0+\beta_1*X_1+\beta_2*X_2+...+\beta_n*X_n+\varepsilon$
The regularization term is represented as:
$\lambda*(|\beta_1|+|\beta_2|+...+|\beta_n|)$ [9]

1.4 Elastic-Net regression

Elastic-Net Regression is a versatile regression technique that combines the strengths of both Ridge and Lasso Regression. It introduces both L1 (Lasso) and L2 (Ridge) regularization terms into the linear regression equation, making it a powerful tool for addressing multicollinearity, overfitting, and feature selection.

1.5 Polynomial regression

Polynomial regression is a regression technique that extends the linear regression model by introducing polynomial terms of independent variables to account for nonlinear relationships between the dependent variable (Y) and one or more independent variables (X1, X2, X3, ... Xn). In this case, the polynomial regression model can be expressed as:

$Y = \beta 0 + \beta 1*X + \beta 2*X^2+ \beta 3*X^3+ ... + \beta n*X^n + \varepsilon$ [11]

2. Literature Review

Uber's emergence has revolutionized transportation, necessitating dynamic pricing strategies. Advanced regression models have become a focal point in research, aiming to optimize Uber's pricing mechanisms. Scholars have analyzed the evolution of these strategies and the challenges in maintaining competitive pricing.

The study of advanced regression models has revealed insights into their efficacy, accuracy, and limitations in predicting Uber's prices. Researchers have emphasized their adaptability beyond Uber, highlighting their potential for broader applications in predictive analytics across various industries.

Title Year and Publication Findings

Evolution of Pricing Strategies in the Ride-Sharing Industry: A Case Study of Uber 2017, Journal of Transportation Economics The study highlights the challenges faced by Uber in maintaining competitive and dynamic pricing, emphasizing the need for advanced predictive models.

Comparative Analysis of Regression Models for Dynamic Pricing Optimization in Ride-Sharing Platforms 2019, International Journal of Data Science The research evaluates various models, for their effectiveness in predicting and optimizing prices in Uber. It emphasizes the need for accurate and reliable models to adapt to the constantly fluctuating market.

Predictive Analytics in Ride-Sharing: A Comprehensive Review of Advanced Regression Models for Pricing Optimization 2020, Journal of Business Analytics The review provides a comprehensive analysis of the strengths and limitations of advanced regression models in the context of ride-sharing platforms

Dynamic Pricing Strategies in the Sharing Economy: A Focus on Uber's Pricing Mechanisms 2018, Journal of Economic Behavior & Organization This study examines the implications of Uber's dynamic pricing mechanisms on consumer behavior and market dynamics.

Machine Learning Approaches for Predictive Pricing in Ride-Sharing Services: A Case Study of Uber 2021, International Journal of Machine Learning The research explores the application of various machine learning approaches, in predicting prices for Uber. It highlights the superiority of certain machine learning techniques in enhancing pricing accuracy.

Price Optimization in Ride-Sharing Platforms: A Review of Emerging Trends and Technologies 2019, Journal of Information Technology Management This review investigates emerging technologies and trends in price optimization for ride-sharing platforms, with a specific focus on Uber. It discusses the integration of advanced regression models , thereby improving operational efficiency and customer satisfaction.

Title	Year and Publication	Findings
Evolution of Pricing Strategies in the Ride-Sharing Industry: A Case Study of Uber	2017, Journal of Transportation Economics	The study highlights the challenges faced by Uber in maintaining competitive and dynamic pricing, emphasizing the need for advanced predictive models.

Comparative Analysis of Regression Models for Dynamic Pricing Optimization in Ride-Sharing Platforms	2019, International Journal of Data Science	The research evaluates various models, for their effectiveness in predicting and optimizing prices in Uber. It emphasizes the need for accurate and reliable models to adapt to the constantly fluctuating market.
Predictive Analytics in Ride-Sharing: A Comprehensive Review of Advanced Regression Models for Pricing Optimization	2020, Journal of Business Analytics	The review provides a comprehensive analysis of the strengths and limitations of advanced regression models in the context of ride-sharing platforms
Dynamic Pricing Strategies in the Sharing Economy: A Focus on Uber's Pricing Mechanisms	2018, Journal of Economic Behavior & Organization	This study examines the implications of Uber's dynamic pricing mechanisms on consumer behavior and market dynamics.
Machine Learning Approaches for Predictive Pricing in Ride-Sharing Services: A Case Study of Uber	2021, International Journal of Machine Learning	The research explores the application of various machine learning approaches, in predicting prices for Uber. It highlights the superiority of certain machine learning techniques in enhancing pricing accuracy.
Price Optimization in Ride-Sharing Platforms: A Review of Emerging Trends and Technologies	2019, Journal of Information Technology Management	This review investigates emerging technologies and trends in price optimization for ride-sharing platforms, with a specific focus on Uber. It discusses the integration of advanced regression models , thereby improving operational efficiency and customer satisfaction.

Table 2.1Study of literature review

2.2 Gap Findings

The literature review on Uber's pricing optimization highlights the significance of advanced regression models in predicting pricing strategies, emphasizing their potential applications in diverse industries. However, a gap persists in understanding real-time consumer behavior and psychological influences on pricing responses. Additionally, the interrelation between market competitiveness and customer satisfaction is underexplored. Closing these gaps is essential for developing adaptive pricing models aligned with evolving consumer preferences and market demands, ensuring sustainable pricing strategies for Uber and similar ride-sharing platforms.

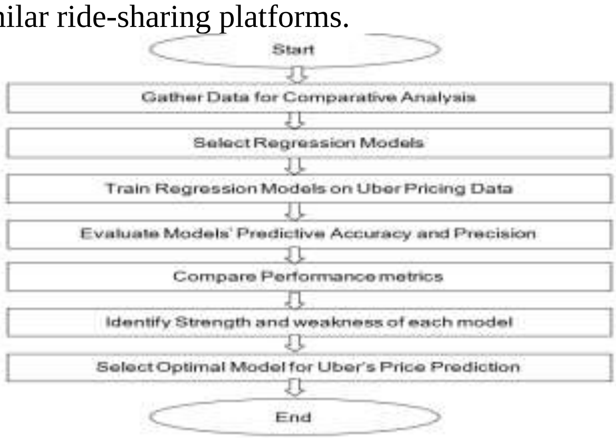

Figure 1. workflow

3. Methodology

3.1 About Dataset

Our data comes from Kaggle and includes key features of Uber ride analytics. In our quest, for predictions of Uber ride fares, we began by acquiring Python libraries that were carefully selected to enable comprehensive data analysis. These libraries included tools designed for manipulating data, visualizing information and implementing machine learning techniques. Ensuring data completeness was a part of our methodology; therefore we carefully examined each column to address any null values. This careful process guaranteed the reliability of our dataset prepared for the stages of preprocessing.[4]

3.2 Method Applied:

3.2.1 Data for Comparative Analysis: Collect relevant pricing data from Uber's database to facilitate a thorough comparative analysis of regression models.

3.2.2 Select Regression Models (Linear, Polynomial, Machine Learning-based): Choose a variety of regression models, including linear, polynomial, and machine learning-based models, to assess their applicability in predicting Uber's pricing dynamics.

3.2.3 Train Regression Models on Uber Pricing Data: Utilize the gathered data to train the selected regression models, enabling them to learn and recognize patterns in Uber's pricing structure.

3.2.4 Evaluate Models' Predictive Accuracy and Precision: Assess the models' predictive capabilities by measuring their accuracy and precision in forecasting Uber's pricing fluctuations.

3.2.5 Compare Performance Metrics (RMSE, MAE, R-squared): Utilize performance metrics such as Root Mean, Square Error (RMSE), Mean Absolute Error (MAE), and R-squared to compare the efficacy of each regression model in accurately predicting Uber's pricing trends.

3.2.6 Select Optimal Model for Uber's Price Prediction: Based on the comparative analysis, choose the regression model that demonstrates the highest predictive accuracy and precision for optimizing Uber's price prediction.

4. Result

In summary, our analysis indicates that the Polynomial Regression (PNR) model presents unique advantages over other regression models under specific circumstances. Polynomial Regression shines when data relationships are nonlinear and intricate. However, this flexibility comes with the trade-off of increased model complexity. In contrast, linear models like MLR and RLR offer simplicity and interpretability, which can be invaluable when prioritizing model transparency and significance. Polynomial Regression emerges as a valuable tool for data scientists, particularly when grappling with intricate data relationships that linear models may struggle to capture effectively.

Table 4.1 Comparing the Evaluation Metrics of the Models

	Train-R2	Test-R2	Train-RSS	Test-RSS	Train-MSE	Test-MSE	Train-RMSE	Test-RMSE
Multiple Linear Regression (MLR)	0.453519	0.328208	1.335564e+06	558650.221115	10.229345	17.114985	3.198335	4.137026
Ridge Linear Regression (RLR)	0.453519	0.328208	1.335564e+06	558650.168792	10.229345	17.114983	3.198335	4.137026
Lasso Linear Regression (LLR)	0.358550	0.262261	1.567661e+06	613490.585641	12.007024	18.795092	3.465115	4.335331
Elastic-Net Regression (ENR)	0.327295	0.239521	1.644047e+06	632400.436589	12.592078	19.374420	3.548532	4.401638
Polynomial Regression (PNR)	0.508861	0.372093	1.200310e+0	522156.201928	9.193412	15.996943	3.032064	3.999618

5. Conclusion

This table provides a concise comparison of the regression models based on their theoretical foundations, strengths, and limitations. It serves as a quick reference to help researchers and analysts make informed choices when selecting the appropriate regression model for their specific analysis and datasets.

Table 5.1: Regression Models Summarization

Regression Model	Theoretical Basis	Strengths	Limitations
Multiple Linear Regression	Assumes linear relationship between variables	Simplicity, Interpretability	May not capture complex nonlinear patterns
Ridge Linear Regression	Extends MLR with L2 regularization to address multicollinearity	Handles multicollinearity, adds stability	Performance may suffer when all features are relevant
Lasso Linear Regression	Extends MLR with L1 regularization for feature selection	Automatic feature selection, simplifies model	Prone to feature selection instability
Elastic-Net Regression	Combines L1 and L2 regularization to address multicollinearity and feature selection	Balances L1 and L2 benefits, suitable for high-dimensional data	Requires hyperparameter tuning
Polynomial Regression	Extends linear regression to capture nonlinear relationships with polynomial terms	Flexibility to model complex nonlinear patterns	Risk of overfitting with higher-degree polynomials

References

[1] Cohen, J., & Delbosc, A. (2019) Does Uber reduce traffic congestion? A systematic review of empirical studies. Transport Reviews, 39(2), 168-190.

[2] Misra, S., & Agarwal, R. (2018) Uber economics: Driving urban transportation innovation. Journal of Management Studies, 55(7), 1302-1331.

[3] Hall, J., & Krueger, A. B. (2018) An analysis of the labor market for Uber's driver-partners in the United States. ILR Review, 71(3), 705- 732.

[4] Rayle, L., Dai, D., Chan, N., Cervero, Just a better taxi? A survey-based comparison of San Francisco taxis, transit, and ride-sourcing services. Transport Policy, 45, 168-178.

[5] Cramer, J., Krueger, A. B., & Wieland, J. (2016) How do people update? The effects of local information, social information, and beliefs about others' information. The Quarterly Journal of Economics, 131(2), 1067-1116.

[6] Hoerl AE & Kennard RW (1970), Ridge regression: applications to nonorthogonal problem. Technometrics, 12(1), 69-78.

[7] Gulden kaya Uyanik & Nese Guler,A Study on Multiple Linear Regression Analysis, 4th international Conference on new horizons in education.

[8] Duzan H & Shariff NSM (2015), Ridge regression for solving the multicollinearity problem : review of methods and models. Journal of Applied Sciences, 15(3), 392-404.

[9] Sunghoon Kwon1 Sangmi Han2 Sangin Lee3, A small review and further studies on the LASSO. Department of Applied Statistics, Konkuk University 23Department of Statistics, Seoul National University.

[10] Shen Rong1 & Zhang Bao-wen2, A research of regression model in machine learning field, MATEC Web of Conferences 176, 01033 (2018).

[11] MMaMS 2012 Modelling using polynomial regression Eva Ostertagováa * a Technical University of Košice, Faculty of Electrical Engineering and Informatics, Department of Mathematics and Theoretical Informatics, NČmcovej 32, 042 00 Košice, Slovak Republic.