Paper 1: Regularization Effectiveness Study

Does Regularization Materially Improve

Generalization on the Bike Sharing Dataset?

Aylin Bozkurt (00161185) and Jana Mumm (00175699)

Faculty of Business, Applied Artificial Intelligence and Digital

Transformation (KDT), Ansbach University of Applied Sciences

Angewandte KI I

Prof. Dr. Sigurd Schacht

November 2, 2025

Confidentiality Statement

I especially confirm that, without any exception, I have fully cited all sources referring to direct quotations of other authors' statements or tables, graphs, and figures, as well as all indirect quotations or modified tables, graphics, and citations from the internet. I am aware that any breach of these rules is considered plagiarism and will be punished according to the general university regulations (Allgemeine Studien- und Prüfungsordnung der Hochschule für angewandte Wissenschaften Ansbach).

During the preparation of this paper, AI-assisted tools such as ChatGPT (GPT-5), Claude Sonnet 4.5, and GitHub Copilot were used in a limited and transparent way. These tools supported code commenting, syntax checking and text clarity improvements. All data analysis, experimental design, model interpretation, and final writing decisions were conducted and verified independently by the authors.

This thesis, in the same or similar form, has not been submitted to any institution yet.
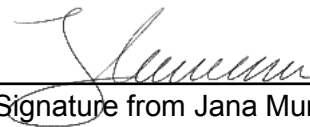
Ansbach, October 30, 2025
_____
Place, Date

_____
Signature from Aylin Bozkurt

_____
Signature from Jana Mumm

Abstract

This study investigates the effectiveness of regularization techniques (Ridge, Lasso, and ElasticNet) compared to standard Linear Regression for predicting hourly bike rentals in Washington D.C.'s Capital Bikeshare system. The research question examines whether regularization materially improves predictive accuracy on the Bike Sharing Dataset (Fanaee-T, 2013), which contains 17,389 hourly observations from 2011 to 2012 with weather, temporal, and environmental features. Preprocessing included one-hot encoding of categorical variables, standardization of numerical features, and an 80/20 train-test split. Models were optimized using five-fold cross-validated grid search to identify optimal regularization strengths and mixing ratios.

Results show all models performed nearly equivalently, with R^2 differences below 0.0001. Statistical tests using Wilcoxon signed-rank and paired t-tests confirmed these differences are not significant (all p-values > 0.50). Regularization provided no measurable improvement over OLS for this dataset. The analysis demonstrates that proper feature engineering, especially one-hot encoding of categorical variables, had a far greater impact, raising R^2 from about 0.39 to 0.68 (a 76% improvement).

These findings suggest that for large, clean datasets with favorable sample-to-feature ratios and low noise, careful preprocessing outweighs regularization benefits. With 17,389 samples and 61 features (ratio 285:1), OLS estimates stable coefficients without penalty-based shrinkage. Regularization may still enhance coefficient stability and interpretability but offers no predictive gains when dataset characteristics favor standard linear regression. Practitioners should therefore prioritize feature engineering over model complexity for similar contexts.

Introduction

Urban bike-sharing systems have been emerging as sustainable transportation solutions in cities worldwide, yet optimizing their operations requires accurate demand forecasting. Predicting bike rental demand is essential for resource allocation, station rebalancing, and service quality enhancement. However, the complexity of factors affecting demand, weather, time of day, seasonality, and user patterns poses challenges for traditional modeling approaches. Linear Regression (OLS) models are attractive for their simplicity and interpretability, but risk overfitting when dealing with several correlated predictors (Géron, 2019, p. 47). Regularization techniques such as Ridge (Hoerl & Kennard, 1970), Lasso (Tibshirani, 1996), and Elastic Net (Zou & Hastie, 2005) address this by penalizing model complexity, shrinking coefficients, and sometimes performing automatic feature selection. As a result, these methods improve model's ability to generalize to unseen data. This paper investigates whether regularization materially improves predictive performance on the Bike Sharing dataset. Specifically, we ask: Does regularization significantly enhance generalization compared to Standard Linear Regression for hourly bike rental prediction? The dataset from the UCI Machine Learning Repository contains 17,389 hourly records from Washington D.C.'s Capital Bikeshare system spanning 2011-2012, including weather, temporal, and environmental features. The task is to predict the total hourly bike rental count (*cnt*). The paper is organized as follows: Section 2 reviews related work on regularization and bike-sharing demand forecasting. Section 3 describes the dataset and preprocessing pipeline. Section 4 outlines the models, hyperparameter tuning strategy, and evaluation metrics. Section 5 presents experimental results with visualizations. Section 6 discusses findings and implications. Section 7 concludes with practical insights and future research directions.

Related Work

Regularization is a cornerstone technique in statistical learning to prevent overfitting and improve model generalization. Ridge regression (Hoerl & Kennard, 1970, pp. 55–67) introduces an L2 penalty to shrink coefficients toward zero. Lasso regression (Tibshirani, 1996, pp. 276–288) adds an L1 penalty that enforces sparsity. ElasticNet (Zou & Hastie, 2005, pp. 301–320) combines both and is especially useful when predictors are correlated. Bike-sharing datasets often contain many interdependent features (e.g., temperature and apparent temperature, or season and month), which can lead to unstable coefficient estimates and poor generalization to new data (Cai et al., 2025; Cortez-Ordoñez et al., 2024). This multicollinearity complicates model fitting and interpretation, highlighting the importance of regularization methods that can mitigate these challenges. Previous studies using bike sharing data applied various approaches such as ARIMA, Random Forests, Gradient Boosting, and LSTMs,

typically using OLS as a baseline (Hamad et al., 2021). However, few systematically compare Ridge, Lasso, and ElasticNet within a linear framework while maintaining interpretability. This work contributes by empirically evaluating whether regularization yields measurable generalization benefits and by analyzing feature importance.

Dataset and Preprocessing

The dataset (Fanaee-T, 2013) includes weather, temporal, and environmental variables recorded hourly in 2011–2012. No missing values were found. Identifiers (*instant*, *dteday*) were removed to avoid artificial temporal ordering. Categorical variables (*season, weathersit, mnth, hr, weekday, yr, holiday, workingday*) were one-hot encoded, resulting in 61 binary features. This step proved crucial: treating them as numeric integers yielded $R^2 \approx 0.39$, while proper one-hot encoding improved performance to ≈ 0.68 (+76%). Numerical features (*temp, atemp, hum, windspeed*) were standardized. The data was split into 80% training and 20% test sets (*random_state = 42*). Exploratory Data Analysis showed that the target variable cnt ranged from 1 to 977, with a mean of approximately 195 and a standard deviation of 182. Hourly usage patterns revealed peaks in the morning and evening corresponding to commuting hours, and seasonal trends showed higher usage in summer months. Temperature and apparent temperature were highly correlated, as were season and month, but these correlations did not affect the stability of OLS estimates.

Methodology

Four linear models were evaluated: Ordinary Least Squares (OLS) as a baseline, Ridge Regression (L2 penalty), Lasso Regression (L1 penalty), and ElasticNet (combination of L1 and L2). Hyperparameters were optimized using five-fold cross-validated grid search. Ridge regression searched 50 logarithmically spaced alpha values from 10^{-3} to 10^3, while Lasso and ElasticNet tested 50 alpha values from 10^{-3} to 10^1. ElasticNet also explored l1_ratio values of 0.1, 0.3, 0.5, 0.7, and 0.9. Five-fold cross-validation ensured robust evaluation across different splits of the training data. Three evaluation metrics were used: R^2, RMSE, and MAE. R^2 measures the proportion of variance explained by the model, RMSE quantifies average prediction error with emphasis on larger deviations, and MAE provides a robust estimate of average error less sensitive to outliers. Together, these metrics give a comprehensive view of predictive performance. Statistical significance of differences between OLS and each regularized model was assessed using paired t-tests and Wilcoxon signed-rank tests with $\alpha = 0.05$, accounting for both parametric and non-parametric assumptions.

Results and Discussion

The performance of all models on the Bike Sharing dataset was very similar. The baseline OLS regression achieved an R^2 of 0.68138 on the test set, with an RMSE of 100.45 and an MAE of 74.11. Ridge, Lasso, and ElasticNet models showed almost identical performance, with differences in R^2 below 0.0002 and minimal changes in RMSE and MAE. Learning curves indicated that training and cross-validation R^2 scores quickly converged as the number of training samples increased, suggesting that the models were neither overfitting nor underfitting. Regularization paths for Lasso and ElasticNet showed that most coefficients remained relatively stable as alpha increased. Only a few coefficients were reduced to zero at larger alpha values, confirming that minimal regularization was required and that more aggressive shrinkage would not improve predictive accuracy. Pairwise statistical comparisons using the Wilcoxon signed-rank test and paired t-test confirmed that none of the regularized models significantly outperformed OLS, with all p-values above 0.50. Small numerical differences are consistent with random variation rather than a true effect of regularization. Feature engineering, particularly the one-hot encoding of categorical variables, had a far greater effect on performance. Correct encoding improved R^2 from roughly 0.39 to 0.68, demonstrating that proper data representation is more important than the choice of regularization technique.

In summary, regularization provided minimal numerical differences without statistical significance. Standard OLS regression is sufficient for well-structured, large-scale datasets. Regularized models remain useful for stabilizing coefficients and improving interpretability, even when predictive performance is largely equivalent to OLS.

Interpretation of Results

These findings suggest that the dataset characteristics strongly influence the impact of regularization. The large sample size of 17,389 observations relative to approximately 61 features allows OLS to estimate stable coefficients without overfitting. The dataset is well-structured and exhibits relatively low noise, which further reduces the need for penalization. Multicollinearity exists between some features, such as temperature and apparent temperature or season and month, but it does not substantially affect predictive accuracy in this context.Feature engineering, particularly the one-hot encoding of categorical variables, had a far greater effect on performance. Proper encoding improved R^2 from roughly 0.39 to 0.68, demonstrating that correct data representation is more critical than the choice of regularization technique for this dataset. Feature selection by Lasso and ElasticNet enhanced interpretability by identifying the most influential variables, such as temperature, apparent

temperature, hour of day, and year. However, these did not result in significant improvements in predictive performance.

In summary, regularization provided minimal numerical differences without statistical significance, highlighting that standard OLS regression is sufficient for well-structured, large-scale datasets. Regularized models remain useful for stabilizing coefficients and interpretability, but their contribution to predictive accuracy in this case is negligible.

Limitations and Future Work

This study is based on a single dataset from Washington D.C. covering the years 2011 and 2012 and focuses exclusively on linear regression models. The results may differ for smaller or noisier datasets where overfitting is more likely and regularization could have a greater effect. Future research could investigate nonlinear or hybrid approaches such as Random Forests, Gradient Boosting, or Neural Networks. It would also be valuable to explore how regularization interacts with feature selection in datasets that are less balanced or have a higher number of features. Temporal modeling techniques, including autoregressive or recurrent models, might capture dynamic usage patterns more effectively and could reveal scenarios in which regularization has a more substantial impact. Even when predictive gains are minimal, regularization can provide more stable coefficient estimates and improve model interpretability, which is valuable in operational contexts.

Conclusion

Regularization techniques including Ridge, Lasso, and ElasticNet did not materially improve predictive performance compared to Ordinary Least Squares regression on the Bike Sharing dataset. Differences in R^2, RMSE, and MAE were minimal, and statistical tests confirmed that these differences are not significant. The main factor influencing predictive performance was careful feature engineering, particularly the one-hot encoding of categorical variables, which increased R^2 from approximately 0.39 to 0.68.

Although regularized models did not enhance predictive accuracy in this case, they provide more stable coefficient estimates and improved interpretability, which can be valuable in operational forecasting. Overall, for large and well-structured datasets, the way data is represented is more important than the choice of regularization method, but regularization remains a useful tool for ensuring robustness and clarity in model interpretation.

Bibliography

Cai, X., Gu, X., Silm, S., Hadachi, A., Jin, T., & Witlox, F. (2025). Differences in bike-sharing usage and its associations with station-surrounding characteristics: A multi-group analysis using machine learning techniques. *Journal of Transport Geography*, *125*, 104201. https://doi.org/10.1016/j.jtrangeo.2025.104201

Cortez-Ordoñez, A., Vázquez, P.-P., & Sanchez-Espigares, J. A. (2024). Features that influence bike sharing demand. *Heliyon*, *10*(18), e37608. https://doi.org/10.1016/j.heliyon.2024.e37608

Fanaee-T, H. (2013). *Bike Sharing* [Dataset]. UCI Machine Learning Repository. https://doi.org/10.24432/C5W894

Géron, A. (2019). *Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems* (Second edition). O'Reilly.

Hamad, S. Y. Y., Ma, T., & Antoniou, C. (2021). Analysis and Prediction of Bikesharing Traffic Flow – Citi Bike, New York. *2021 7th International Conference on Models and Technologies for Intelligent Transportation Systems (MT-ITS)*, 1–8. https://doi.org/10.1109/MT-ITS49943.2021.9529290

Hoerl, A. E., & Kennard, R. W. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, *12*(1), 55–67. https://doi.org/10.1080/00401706.1970.10488634

Tibshirani, R. (1996). Regression Shrinkage and Selection Via the Lasso. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *58*(1), 267–288. https://doi.org/10.1111/j.2517-6161.1996.tb02080.x

Zou, H., & Hastie, T. (2005). Regularization and Variable Selection Via the Elastic Net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *67*(2), 301–320. https://doi.org/10.1111/j.1467-9868.2005.00503.x