**STT 805**

**STT Modeling for Business Analytics**

**Project Report**

**Demand Forecasting for a Bike Sharing platform**

A close up of text on a black background

Description automatically generated

**TEAM:**

Sharan Salem - salemsh1@msu.edu

Karan Dalal - dalalkar@msu.edu

Neil Joshi - joshine2@msu.edu

Syed Kashif Mujtaba - kamoonpu@msu.edu

Vishal Agarwal - agarwa97@msu.edu

1. **Abstract**

Inventory management and budget planning are one of the major decision areas for managers of any company. In the era of data analytics, such decisions are driven by demand forecasting. For the context of this project, we are forecasting user demand for a bike-sharing platform. Existing research has used a vast variety of statistical techniques like Neural Networks, Random Forest, and Polynomial regression to forecast demand. Additionally, in actual practice, the Linear regression model is considered a very dependable statistical method for predicting a quantitative variable. The current study on this dataset shows that some studies have also used linear regression techniques. However, for this existing research, the authors have not considered validating linear regression assumptions while making the best fit model to forecast demand.

To fill up this gap, our project incorporates Linear regression to forecast daily demand for the number of bikes that can be rented out for a Washington based bike-sharing company considering the assumptions discussed earlier. The factors that determine the demand include weather-related parameters and day/time-related parameters. While making a linear regression model, the assumptions for the best fit were one of the most important considerations for model creation. Feature selection techniques and regularization approach are also applied, and finally, the model selection was based on error comparison and satisfactory acceptance of linear regression assumptions. Advance concept like predictor attribute transformation are also considered for best model creation. In conclusion, the model deduced using regsubset feature selection is chosen as the final predictive model for suggesting to the company managers.

1. **Motivation**

An online bike-sharing platform is a new age bike rental system where the whole process of membership, rental, and return is automatic. The system facilitates a user to easily rent a bike from a location and return it to another location. Currently, bike-sharing companies are very prominent in China and are picking up pace in the USA. They command high private equity capital from investors. In recent times, using bikes as a mode of commutation serves as an important characteristic in traffic, environmental, and health issues

The project aims to help forecast the demand for the bike-sharing company based on the number of available attributes given in the data. This should help them to create an inventory and workforce-related strategy apart from financial and maintenance estimates. Also, because of the improving bike-sharing culture, the project can have an indirect impact on promoting health, environment, and other social impacts on society.

1. **Research Methodology**
2. **Attribute Description**

• instant: record index

• dteday : date in format of yyyy-mm-dd

• season : season (1:Jan-Mar, 2: Apr-June, 3:July-Sep, 4:Oct-Dec)

• yr : year of operation (0: 2011, 1:2012)

• mnth : month (1 to 12, starting January)

• holiday : whether a particular day is a holiday or not \*

• weekday : particular day of the week, 0 is Sunday

• workingday : if day is neither weekend nor holiday is 1, otherwise is 0.

• weathersit :

1: Clear, Few clouds, Partly cloudy, Partly cloudy

2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist

3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds

4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog

• temp : Actual temperature in Fahrenheit\*\*

• atemp: Actual feeling temperature in Fahrenheit\*\*

• hum: Actual humidity \*\*

• windspeed: Actual wind speed in mph \*\*

• casual: count of casual users

• registered: count of registered users

• cnt: count of total rental bikes including both casual and registered

*\* data has been collected from dchr.dc.gov/page/holiday-schedule*

*\*\* data has been collected from wonderground.com/history*

1. Results
2. Conclusion
3. Future Scope

* Currently, the demand forecasting was done on daily basis. As we hourly demand data also available on UCI ML data repository, we can perform hourly demand prediction as well.
* Attributes having similar coefficients can be clubbed together and considered as single feature. This help in reducing number of features and decreasing model complexity.
* Given the right data, the scope of the study can be increased other regions to gather more data points or to test model accuracy.
* Exploration of other feature selection methods like forward and backward stepwise can also be incorporated.

1. Bibliography
2. Reference Book: Kutner, Nachtsheim, Neter: Applied Linear Regression Models, 4th Edition
3. Data Source: https://archive.ics.uci.edu/ml/datasets/bike+sharing+dataset
4. Temperature data: www.wunderground.com/history
5. Capital Bikeshare: https://www.capitalbikeshare.com/
6. Linear Regression: https://en.wikipedia.org/wiki/Linear\_regression
7. Linear Regression using R: http://r-statistics.co/Linear-Regression.html
8. GGPLOT2 Cheat Sheet: https://rstudio.com/wp-content/uploads/2015/03/ggplot2-cheatsheet.pdf
9. RegSubsets: https://cran.r-project.org/web/packages/leaps/leaps.pdf
10. Cross Validation: http://www.sthda.com/english/articles/38-regression-model-validation/157-cross-validation-essentials-in-r/
11. Coefficient and P-value interpretation: https://statisticsbyjim.com/regression/interpret-coefficients-p-values-regression/
12. Rainbow Test: http://math.furman.edu/~dcs/courses/math47/R/library/lmtest/html /raintest.html
13. Jarque-Bera Test For Normality: https://www.rdocumentation.org/packages/tsoutliers/versions /0.3/topics/jarque.bera.test
14. Breusch-Pagan Test: https://www.rdocumentation.org/packages/lmtest/versions/0.9-37/topics/bptest
15. Goldfeld-Quandt test against heteroskedasticity: http://math.furman.edu/~dcs/courses /math47/R/library/lmtest/html/gqtest.html
16. Performs the Durbin-Watson test for autocorrelation of disturbances: http://math.furman.edu /~dcs/courses/math47/R/library/lmtest/html/dwtest.html