

# Online News Popularity: Study of Global Subjectivity impact on Engagement through Shares

Ruchit Patel  
Pace University  
[rp90622n@pace.edu](mailto:rp90622n@pace.edu)  
GitHub - <https://github.com/Ruchit42>

# Project Proposal



- Targeted problem
  - I was shocked when I found out that according to an article on [pewresearch.org](http://pewresearch.org)
    - 48% of U.S adults say they get news from social media “often”
    - Out of all the social media, Facebook outpaces all other social media sites as nearly a third of Americans regularly get news on Facebook
  - After coming across this facts, I wanted to know what and how online news spreads and more specially, which type of stories are shared more than other.
- Research question
  - How does Global Subjectivity impact the number of news Shares on social media? (News that are shared the most are subjective or Objective?)
  - Can we build a machine learning model which can predict the number of shares an article is going to get bases on the features on the articles?
- Dataset
  - <https://archive.ics.uci.edu/dataset/332/online+news+popularity>
  - The articles were published by Mashable ([www.mashable.com](http://www.mashable.com)), which holds the rights to the content and its replication. As a result, only a subset of the statistics pertaining to the original content are shared in this collection.
  - #rows - 39,644
  - Parameters - 61
- Motivation
  - The challenge is to determine why some news stories are shared more frequently than others. Understanding what makes one news item more popular than others is crucial given that more and more people are obtaining their news through social media and apps like Reddit and WhatsApp.

# Predicating redicating reservoir sensitivity rapidly with single-correlation analysis and multiple regression

Y. Sun, C. Xiao and Y. Lang, "Predicating reservoir sensitivity rapidly with single-correlation analysis and multiple regression," 2011 6th IEEE Joint International Information Technology and Artificial Intelligence Conference, Chongqing, China, 2011, pp. 100-104, doi: 10.1109/ITAIC.2011.6030286.

**Goal:** Using single-correlation analysis and multiple regression to predict reservoir sensitivity

**Dataset:** Reservoir water sensitivity

**Methodology:**

- Single-Correlation Analysis
  - Measure the linearity between two variables
- Maintaining the Integrity of the Specification
  - MLR Model

**Results:** The combining method is an ideal new technique to forecast future reservoir damage sensitivity and can provide a solid basis for building strategies to preserve reservoirs.

# Predicting Stock Price Movements Based on Different Categories of News Articles

Y. Shynkevich, T. M. McGinnity, S. Coleman and A. Belatreche, "Predicting Stock Price Movements Based on Different Categories of News Articles," 2015 IEEE Symposium Series on Computational Intelligence, Cape Town, South Africa, 2015, pp. 703-710, doi: 10.1109/SSCI.2015.107.

**Goal:** This study investigates how the simultaneous usage of news pieces with varying degrees of relevance to the target stock can enhance financial forecasting outcomes.

**Dataset:** Stock-Price Dataset

**Methodology:** Two approaches

- The SVM and KNN Approaches
  - System is trained on different GICS classification levels
- The MKL Approach
  - The following kernel combinations were taken into consideration: two Gaussian, two linear, and two polynomial kernels; two Gaussian and two linear; two Gaussian and two polynomial; two linear and two polynomial; and finally, two Gaussian, two linear, and two polynomial kernels combined.

**Results:** In comparison to algorithms relying on a single news category, MKL outperformed SVM and kNN algorithms based on specific news kinds (SS, SIS, IS, GIS, or SeS), demonstrating superior outcomes in terms of trading returns and accuracy.

# The Electricity Price Prediction of Victoria City Based on Various Regression Algorithms

S. Orenc, E. Acar and M. S. Özerdem, "The Electricity Price Prediction of Victoria City Based on Various Regression Algorithms," 2022 Global Energy Conference (GEC), Batman, Turkey, 2022, pp. 164-167, doi: 10.1109/GEC55014.2022.9986605.

**Goal:** Predicting the price of electricity with the help of four models to compare which model will prove to be more accurate

**Dataset:** Daily Electricity Price and Demand Data announced

**Methodology:** Created and compared the result from four regression model

- Linear Regression
- Decision Tree Regression
- Gradient Boosting Regression
- Random Forest Regression

**Results:** gradient boosting is the best algorithm to predict electricity price according to the study with the MAE score of 0.49

# Research on Prediction of Traffic Flow at Non-detector Intersections Based on Ridge Trace and Fuzzy Linear Regression Analysis

Y. Liu and M. Sha, "Research on Prediction of Traffic Flow at Non-detector Intersections Based on Ridge Trace and Fuzzy Linear Regression Analysis," 2009 International Conference on Computational Intelligence and Security, Beijing, China, 2009, pp. 571-575, doi: 10.1109/CIS.2009.35.

**Goal:** measure and predict traffic flow about 19 different junctions in Australian city

**Dataset:** Traffic Flow Intersection

**Methodology:**

- Used Ridge Regression to predict traffic flow at a certain intersection at certain time of the day

**Results:** The experiment's findings demonstrate that it is possible to utilize the FLR model, which uses ridge regression analysis as its foundation, to estimate traffic flow at crossings without detectors. This model uses accurate data as its input and symmetric triangular fuzzy numbers as its output.

---

Article Link - <https://ieeexplore.ieee.org/document/5376191>

# Sales Prediction using Online Sentiment with Regression Model

S. K. Punjabi, V. Shetty, S. Pranav and A. Yadav, "Sales Prediction using Online Sentiment with Regression Model," 2020 4th International Conference on Intelligent Computing and Control Systems (ICICCS), Madurai, India, 2020, pp. 209-212, doi: 10.1109/ICICCS48265.2020.9120936.

**Goal:** The main goal of this essay is to forecast a car's sales using sentiment analysis from numerous online sources.

**Dataset:** Car Sales Volume

**Methodology:**

- Polynomial Regression
- Feature engineering using sentiment weightage

**Results:** Polynomial degree 2 and 3 r2score of 0.9005 and 0.9016, respectively. Calculating the degree of the polynomial was necessary to resolve overfitting because it involved comparing all of the degrees of the 100% fitted data, together with the train and test split.

# A Comparison of Regression Techniques for Prediction of Air Quality in Smart Cities

K. D. Garg, M. Gupta, B. Sharma and I. B. Dhaou, "A Comparison of Regression Techniques for Prediction of Air Quality in Smart Cities," 2023 1st International Conference on Advanced Innovations in Smart Cities (ICAISC), Jeddah, Saudi Arabia, 2023, pp. 1-6, doi: 10.1109/ICAISC56366.2023.10085369.

**Goal:** To evaluate the accuracy of the air quality index (AQI) projection of PM 2.5 in Chandigarh, India, this study compares and contrasts the effectiveness of various machine learning techniques.

**Dataset:** Climate Data set and AQI Data Set

**Methodology:**

- Regression
  - Calculating Mean Squared Error, Mean Absolute Error and Root Mean Squared Error to measure accuracy
  - Lasso Regression Model and Random Forest Model

**Results:** Compared to the other three algorithms in the smart city, the random forest algorithm predicts PM 2.5 more accurately.



# Generative Adversarial Network for Robust Regression using Continuous Dataset

Y. -L. Min, S. -J. Hong, H. -j. Kim and S. -I. Lee,  
"Generative Adversarial Network for Robust  
Regression using Continuous Dataset," 2020  
International Conference on Information and  
Communication Technology Convergence (ICTC),  
Jeju, Korea (South), 2020, pp. 1209-1211, doi:  
10.1109/ICTC49870.2020.9289188.

**Goal:** To design training architecture for robust regression. Using nonlinear regression to solve limitation of linear regression

**Dataset:** –

**Methodology:**

- Applying adversarial architecture of GAN to regression problem
- Use adversarial architecture to increase the performance of the discriminator which performs as regressor

**Results:** the model trained with the proposed architecture can show the better performance on regression datasets. It is considered that applying GAN to regression model is worthy of further exploration for solving regression problems.

---

Article Link - <https://ieeexplore.ieee.org/document/9289188>

# Comparison of Linear Regression and Logistic Regression Algorithms for Ground Water Level Detection with Improved Accuracy

C. G. Raju, V. Amudha and S. G, "Comparison of Linear Regression and Logistic Regression Algorithms for Ground Water Level Detection with Improved Accuracy," 2023 Eighth International Conference on Science Technology Engineering and Mathematics (ICONSTEM), Chennai, India, 2023, pp. 1-6, doi: 10.1109/ICONSTEM56934.2023.10142495.

**Goal:** Improve the accuracy of both Linear Regression and Logistic Regression algorithms for Ground Water Level Detection

**Dataset:**Ground water Detection

**Methodology:** Participants were divided into two groups and assessed using the Novel Linear Regression method and the Logistic Regression Algorithm to predict groundwater levels. The Novel Linear Regression method exhibited a high accuracy rate of 93.23%, while the Logistic Regression Algorithm achieved an accuracy rate of 86.5%. Both groups consisted of 15 samples each, totaling 30 instances of accurate predictions.

**Results:**The Accuracy of the Novel Linear Regression algorithm is 93% and the accuracy of the Logistic Regression algorithm is 85%. Novel Linear Regression is the method of choice.

**Article Link -**

<https://ieeexplore.ieee.org/document/10142495/metrics#metrics>