

Predicting Virality with eXtreme Gradient Boosting on Online News Popularity Data

Ednalyn C. De Dios

Western Governors University

September 16, 2023





Ednalyn C. De Dios

Preferred Nickname: “Dd”

Machine Learning Engineer at Microsoft

US Navy Veteran



5 Steps to Data Science Success

1. Planning
2. Getting the Data
3. Preparing the Data
4. Exploratory Data Analysis
5. Modeling



Planning







Can gradient boosting be constructed based solely on the Online News Popularity Data?

- H_0 : Gradient boosting cannot be made from the Online News Popularity dataset.
- H_1 : Gradient boosting can be made from the Online News Popularity dataset.



Getting the Data





UC Irvine
Machine Learning
Repository

Online News Popularity

39,979 rows

61 columns

Mashable.com



Preparing the Data



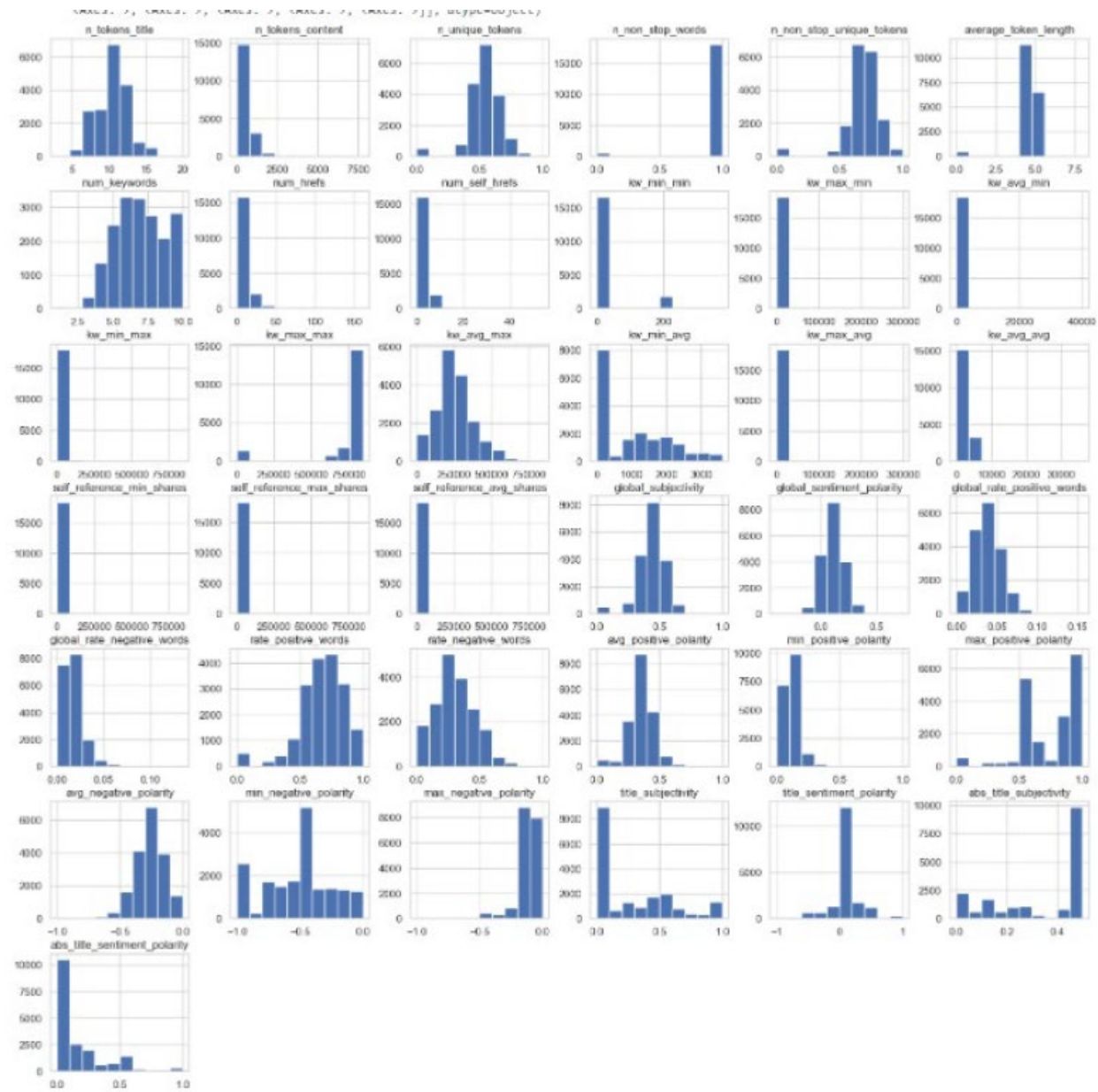


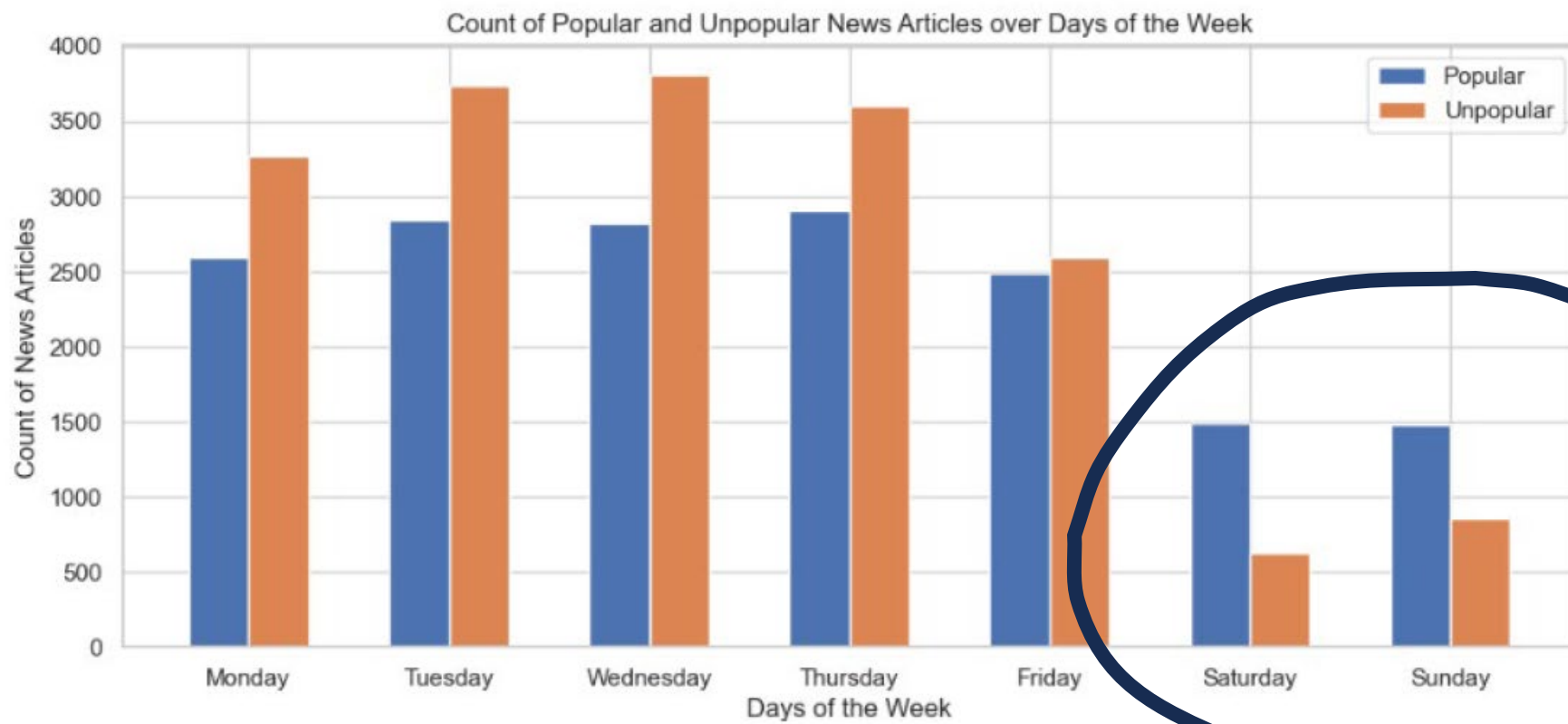
-
- Check for missing values
 - Correct the column names
 - Drop duplicate entries
 - Create the target variable
 - Remove outliers



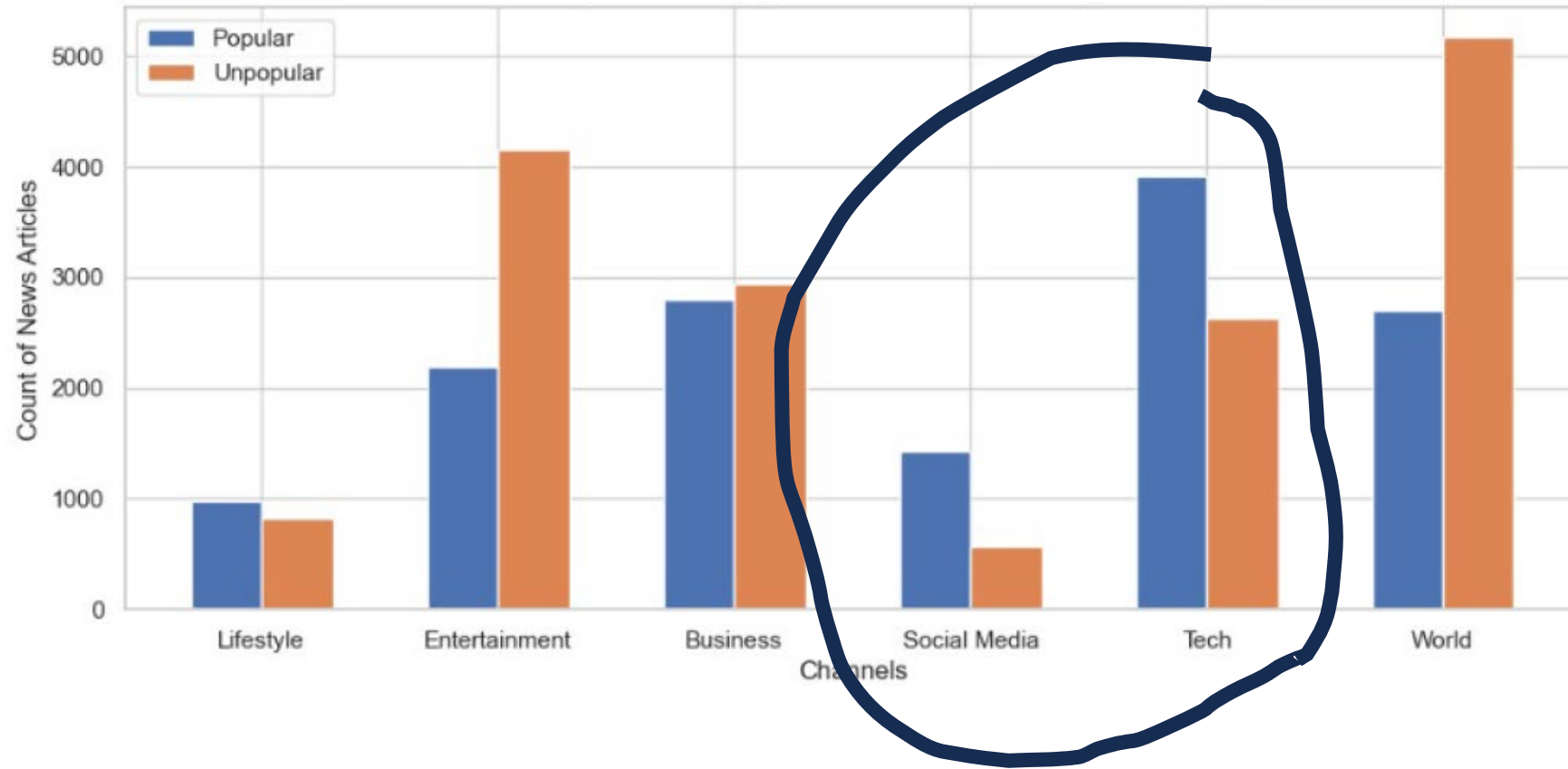


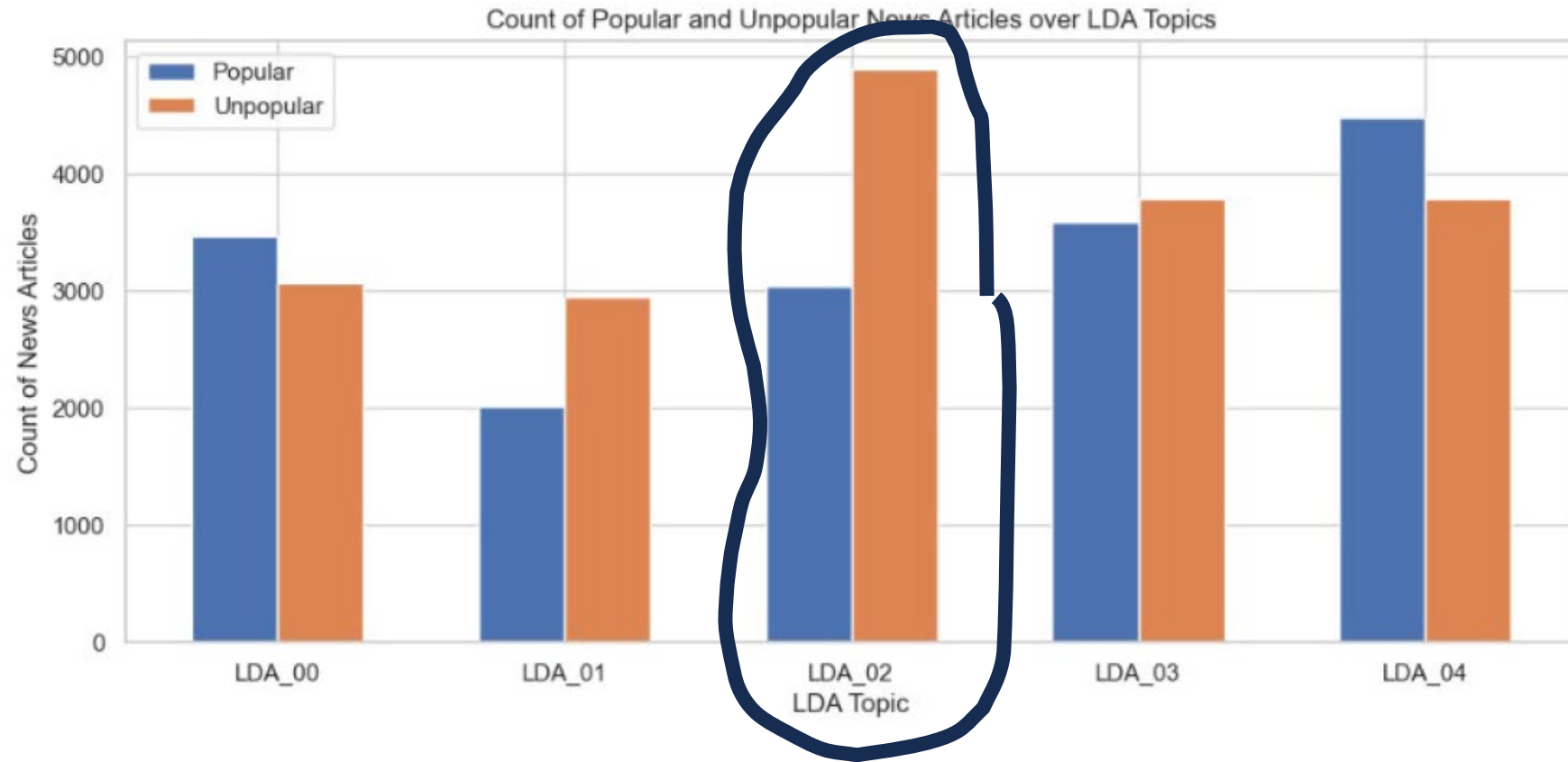
Exploratory Data Analysis





Popular Vs. Unpopular Channels

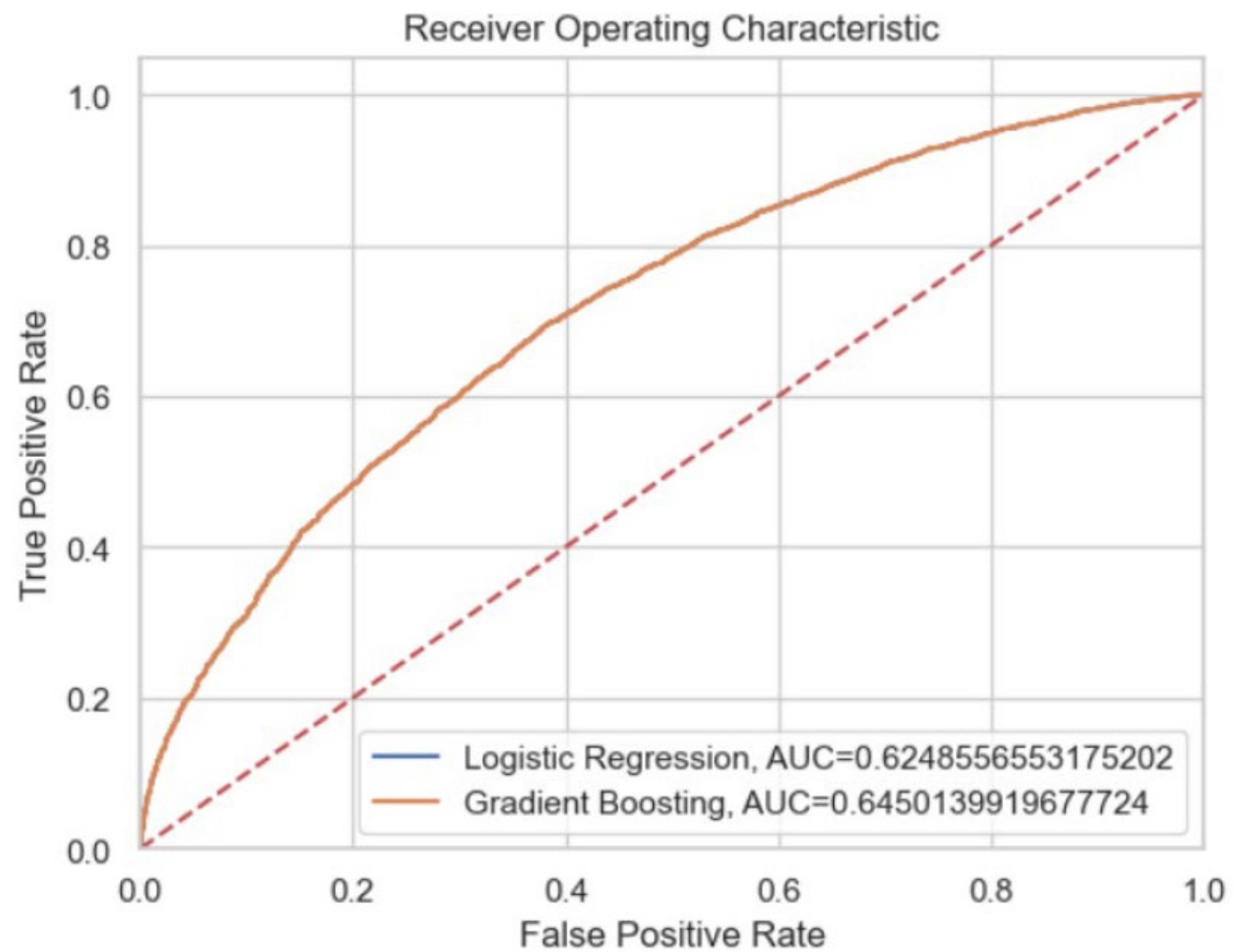






Modeling









Limitations



Niched nature of the articles



Proposed Actions



01

Discern what type
of articles to
publish on the
weekend.

02

Use XGBoost
regression on the
number of shares.

03

Use clustering.



Benefits



With feature extraction, content creators can optimize their articles.



Fin.