A Comprehensive Review on Online News Popularity Prediction using Machine Learning Approach

SVM [4]	No. of features {1,5,10,20}	Average accuracies were above 60% in 8 out of 10 cases. For new York times and Seattle times above 70%
AdaBoost [1]	No. of tress {10,20,50,100,200,400}	AUC = 72%
KNN [1]	No. of neighbors {1,3,5,10,20}	AUC = 67%
Naïve Bayes [1]	Metrics computed over the union of all 29 test sets.	AUC = 65%
Linear Regression [2]	Target values were discretized to binary categories	Accuracy = 66%
Logistic Regression [11]	Data is classified and stochastic gradient ascent rule was used to implement it.	Accuracy = 66%
Genetic Algorithm [12]	Used genetic algorithm to get the optimum attributes and further classified the data using naïve bayes.	Accuracy =91.96%

My Results

Model	Mean Squared Error	R-squared
Linear Regression	0.054613	0.172152
Decision Tree Regression	0.075507	-0.144565
Random Forest Regression	0.057817	0.123579
SVR	0.061173	0.072712
Gradient Boosting Regression	0.056502	0.143513
Neural Network Regression	47913.430737	-726291.346984

SVM: 7%

Linear Regression: 17%

Analysis of book sales prediction at Amazon marketplace in India: a machine learning approach

Table 3 Regression results (MS Excel output)

Regression statistics		
Multiple R		0.708055
R square		0.501342
Adjusted R square		0.499207
Standard error		1.357702
Observations		1408
	Coefficients	*P* value
Intercept	9.122841	1.10E−57
ln(curr_price)	0.643128	*6.20E−14*
ln(disc_amt)	−0.15923	0.088075567
ln(disc_rate)	0.236044	0.07627463
ln(review_volume)	0.652	*5.58E−147*
ln(nop)	0.33616	*5.24E−07*
average_ratings	0.02444	0.663777605

Table 5 Regression results (MS Excel output)

Regression statistics		
Multiple R		0.714676848
R square		0.510762997
Adjusted R square		0.507613402
Standard error		1.346258511
Observations		1408
	Coefficients	*P* value
pos_senti	0.198632539	*0.006329751*
neg_senti	−0.269830938	*0.000667931*
overall_polarity	−0.188711007	*0.029535947*

Table 6 Regression results (MS Excel output)

Regression statistics		
Multiple R		0.723079238
R square		0.522843585
Adjusted R square		0.517330168
Standard error		1.331917191
Observations		1408
	Coefficients	*P* value
disc * rev	−0.600198927	*1.3E−102*
disc * avg	0.055090136	0.783823
disc * pos_senti	−0.004198911	0.070894
disc * neg_senti	−0.003995277	0.304932
rev_vol * pos_senti	−0.000955085	*0.000262*
rev_vol * neg_senti	−0.001445835	*0.003966*

My Results

```
Linear Regression:
Mean Squared Error: 0.0546130244116619 16
R-squared: 0.17215233670751073
```

Linear Regression: 17%

Machine Learning Methods for Online News Popularity Prediction

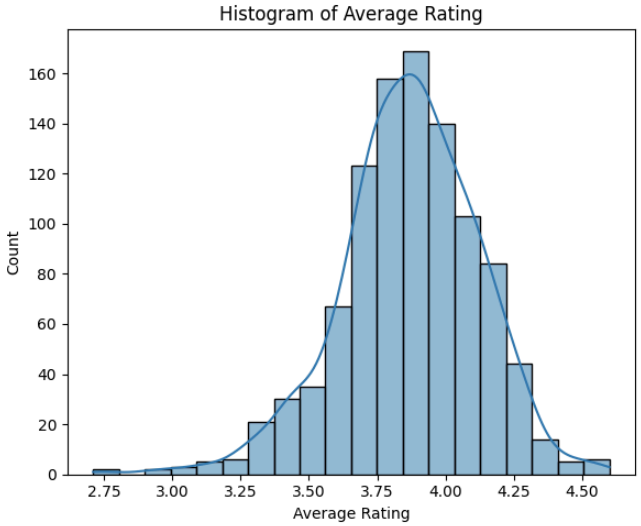
TABLE 2: ACCURACY

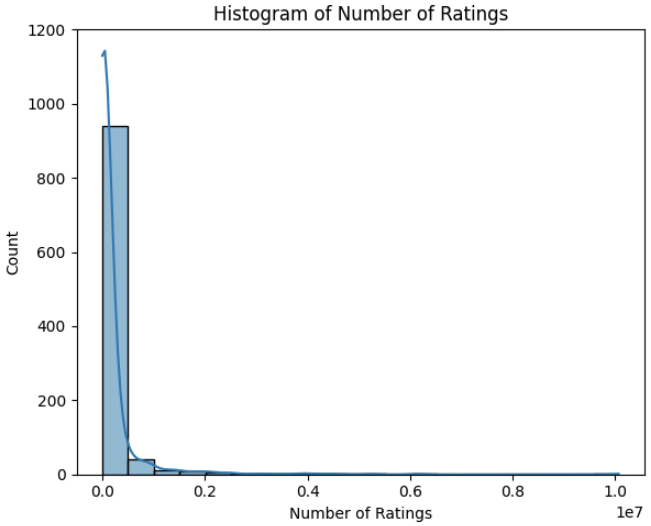
Algorithm	Decision Tree	KNN	Random	Naïve Bayes
Accuracy (Original)	57.9%	57%	66.4%	61.2%
Accuracy (Cross Validation)	55.5%	54.9%	64.4%	60.8%

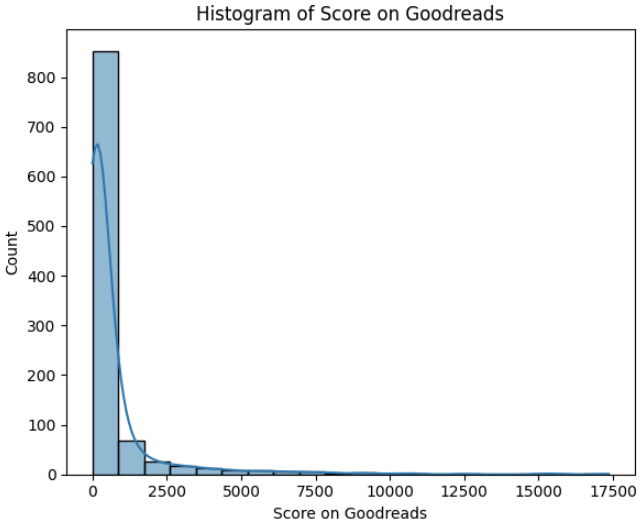
My Results

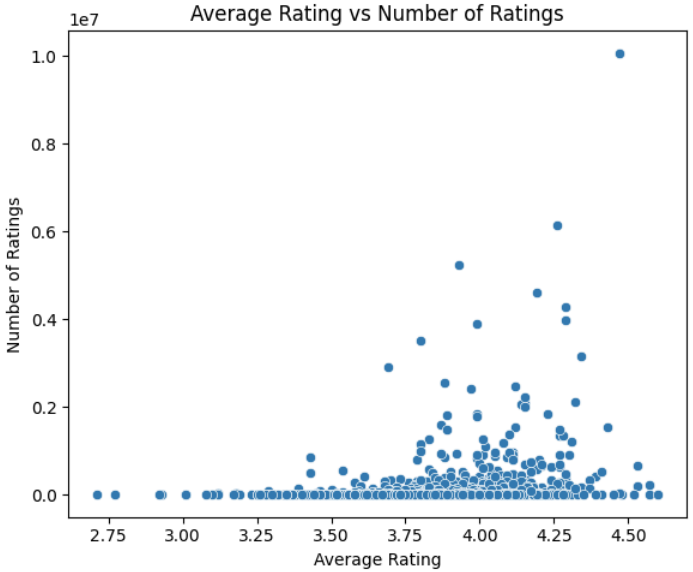
```
Decision Tree:
Mean Squared Error: 0.11092192723697147
Standard Deviation: 0.043004362272931314
```

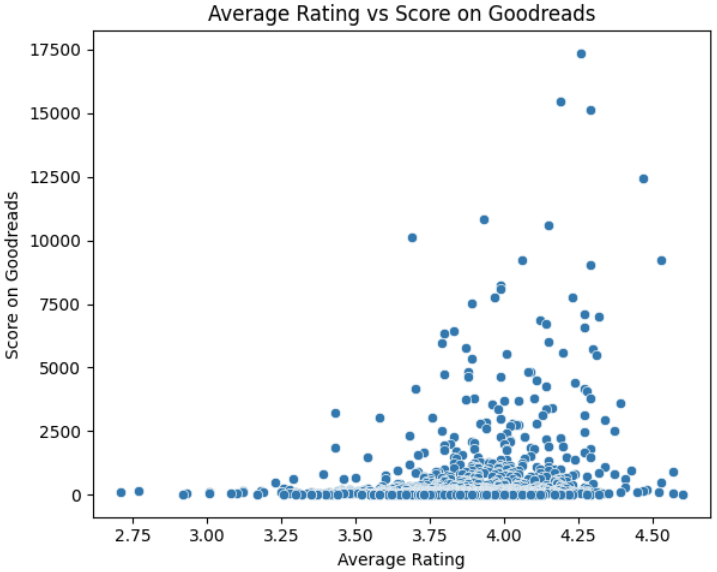
My Visualizations











References

Rathord, P., Jain, A., & Agrawal, C. (2019). A comprehensive review on online news popularity prediction using machine learning approach. trees, 10(20), 50.

Sharma, S. K., Chakraborti, S., & Jha, T. (2019). Analysis of book sales prediction at Amazon marketplace in India: a machine learning approach. Information Systems and e-Business Management, 17(2), 261-284.

Sitapara, R., Kotian, P., Chaudhary, P., & Kamble, S. (2018). Machine learning methods for online news popularity prediction. In IC-CSOD-2018 Conference Proceedings (p. 309).