Predicting UMD Professor Ratings

Varun Nautiyal

Overview

- PlanetTerp API provides course & professor data at UMD
- Goal: predict average professor rating from student review text (no direct student ratings), utilizing sentiment analysis
- Three unique models: Linear Regression, Random Forest, KNN
- This may be interesting because:
  - It can help students pick the best instructors before registration
  - It turns unstructured review text into actionable insights
  - It can scale to other campuses or course evaluation systems

Data Collection

- Fetched all UMD professors via *planetterp* wrapper
- Flattened approximately 38,000 reviews from 4,500 professors into a DataFrame
- Utilized threads to optimize data retrieval process

PlanetTerp includes...

24,097 courses

13,427 professors

37,554 reviews

323,979 course grades

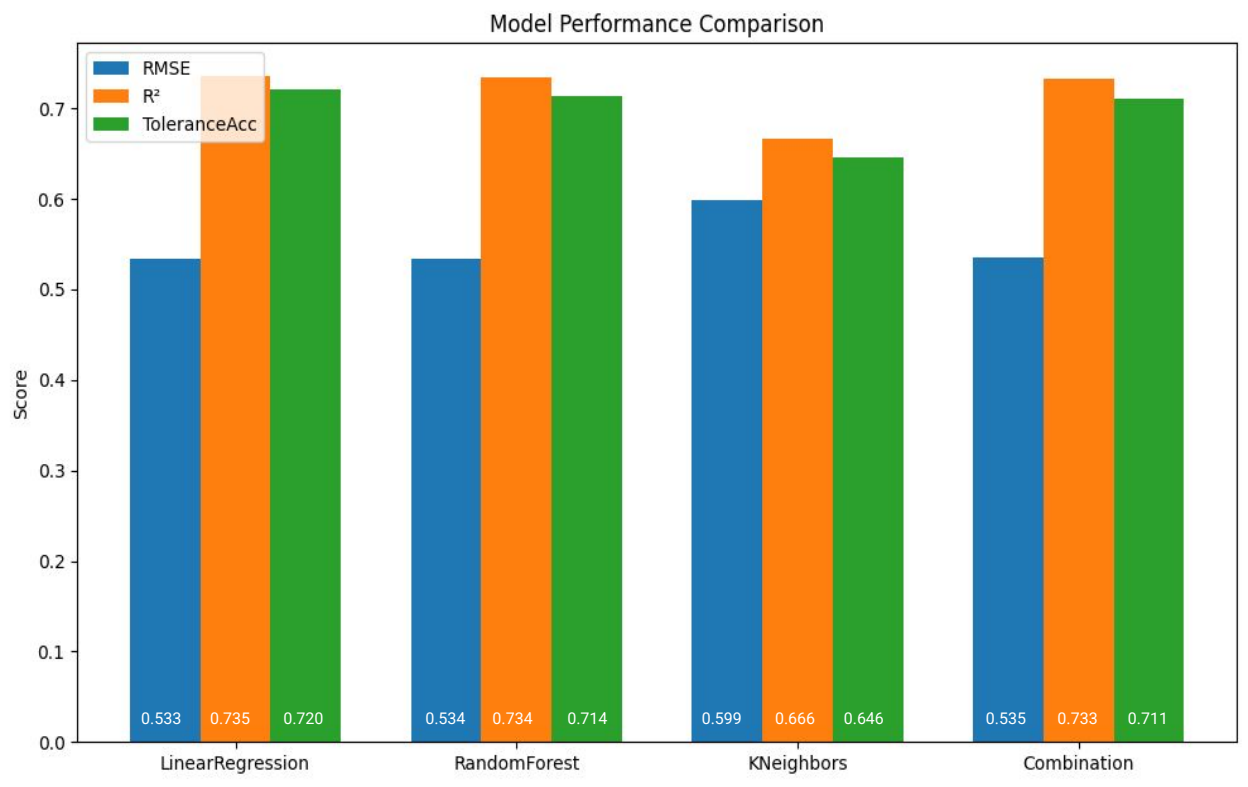
Feature Engineering

- Sentiment scored with VADER (positive, negative, or neutral)
- Aggregated per-professor:
  - *pos_count, neg_count, neu_count, total, pos_ratio*
- Filtered out professors with fewer than 6 reviews → 1,600 professors remain
  - Tuned to obtain optimal model accuracy

Training/Testing

- Features: the five sentiment aggregates
- Target: average_rating from PlanetTerp
- 80/20 train/test split (random_state = 42 for reproducibility)
- Models:
  - Linear Regression (no scaling needed)
  - Random Forest: 100 trees, max_depth = 10
  - KNN: k = 5, manually standardized features
  - Combined: average of the outputs from the three models

Results


Model Performance Comparison
RMSE
R²
ToleranceAcc
Score
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.533
0.735
0.720
0.534
0.734
0.714
0.599
0.666
0.646
0.535
0.733
0.711
LinearRegression
RandomForest
KNeighbors
Combination

Model Performance

Rankings:

1. Linear Regression had best overall fit; RMSE (0.533) was the lowest and R^2 (0.735) was the highest
2. Combined average smoothed out each model's quirks, but still was slightly worse than LR (RMSE: 0.535, R^2: 0.733)
3. Random Forest nearly matched LR on R^2 (0.734) but slightly higher RMSE (0.534)
4. K-Nearest Neighbors trailed behind, with a noticeably higher RMSE (0.599) and lower R^2 (0.666)

Accuracy was highest for LR (0.720), followed by RF (0.714), combined (0.711), and KNN (0.646)

Conclusions

- High success: Achieved up to $R^2 = 0.74$, RMSE = 0.53, and 72% of predictions within ±0.5 stars
  - Nearly ¾ of professors predicted to within half a star, a practically acceptable margin for course decisions
  - Tolerance measured on the held-out 20% test set, reflecting true generalization

Evaluation Process

- RMSE: shows average star-rating deviation, with closer to 0 indicating predictions are closer to actual ratings
- R^2: measures how much of the rating variability the model captures where higher means stronger explanatory power
- ±0.5-star tolerance accuracy: reflects practical success; percent of professors predicted “close enough” (within half a star) for real world applications

Important Features

- Since this model relied solely on sentiment analysis for each review, there were no additional features for comparison, but there are some components which may be important to observe:
  - Positive-to-Total Ratio: strongest single correlate of higher ratings
  - Total Review Count: represents professor visibility/exposure
  - Positive & Negative Counts: raw sentiment volumes add context