

A decorative graphic on the left side of the slide consisting of two overlapping parallelograms. The front one is blue and the back one is a light mint green. They are positioned diagonally, with the blue one partially covering the green one.

Wiki Learn

an interactive curriculum developer
for the modern web

by Quinn Dizon



Overview

What is it?:

- Part 1: Recommender System
 - Part 2: ML Classification model
 - Accessible via web app
-

Data:

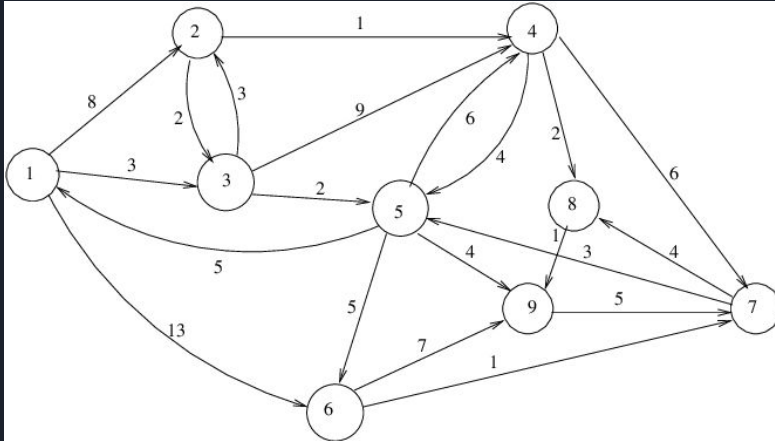
- Wikipedia API
 - Data retrieved in real time
 - Always current with newest edits
 - No limit on topics

Who is it for:

- Individuals:
 - Quickly learn about the topics associated with a discipline
 - Circumvent the *Google black hole*
- Education & Business:
 - Help develop course/training curriculum
 - Find knowledge holes in existing curricula

Part 1 - Recommender System

Nodes = articles | Edges = links to/from



Features:

- Network Connectivity:
 - Degrees
 - in/out edges
 - Centrality
 - Page rank
 - Reciprocity
- Relation to Entry Article
 - Shared categories
 - Shared neighbors
 - Jaccard similarity
 - Shortest paths to/from



Collaborative Filtering: Item to Item

Feature Division:

- Bonuses
 - Shared categories with entry
 - Shared neighbors with entry
 - Centrality
 - Page rank
 - Reciprocity
- Penalties
 - Path length to/from entry
 - Degrees (/ mean degrees)

Similarity Rank

$$\frac{b_1 + b_2 \dots + b_n}{p_1 + p_2 \dots + p_n}$$

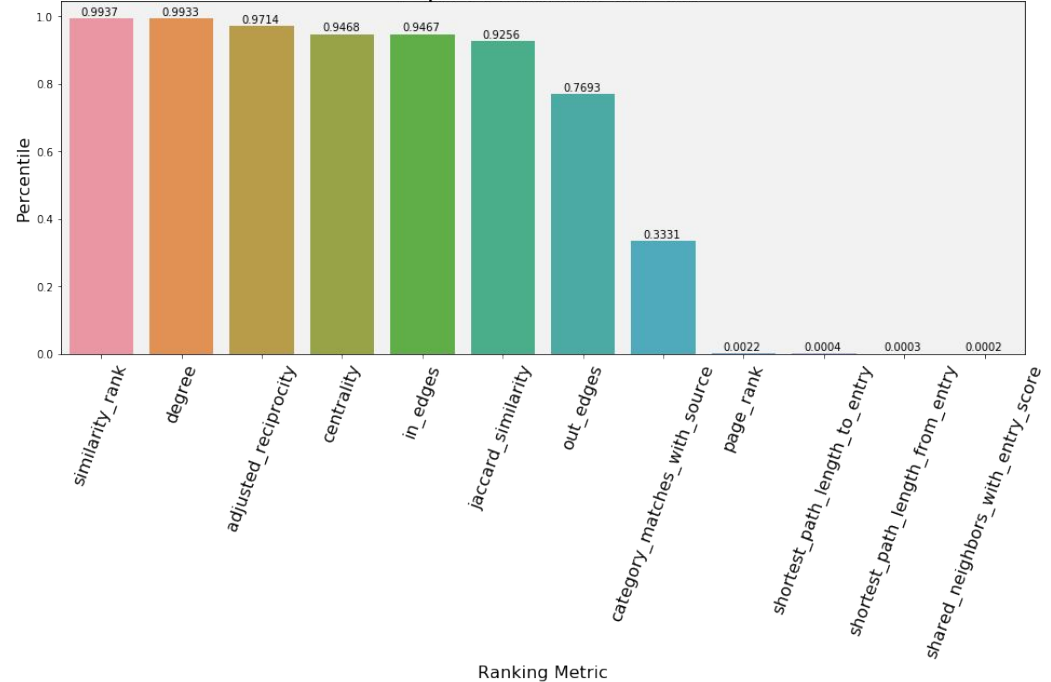
Result Validation

- 7 RF in scientific works
- 8 Open source implementations
- 9 **See also**
- 10 References
- 11 Further reading
- 12 External links

See also [\[edit \]](#)

- [Boosting](#)
- [Decision tree learning](#)
- [Ensemble learning](#)
- [Gradient boosting](#)
- [Non-parametric statistics](#)
- [Randomized algorithm](#)

Top Percentile to Capture all 'See Also' Links
Topic: Random forest



Part 2 - Classification Model

Classes:

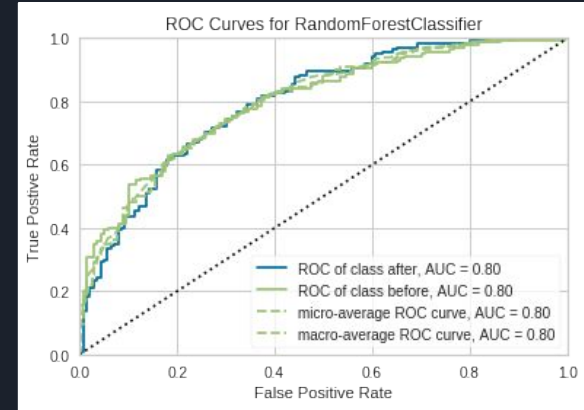
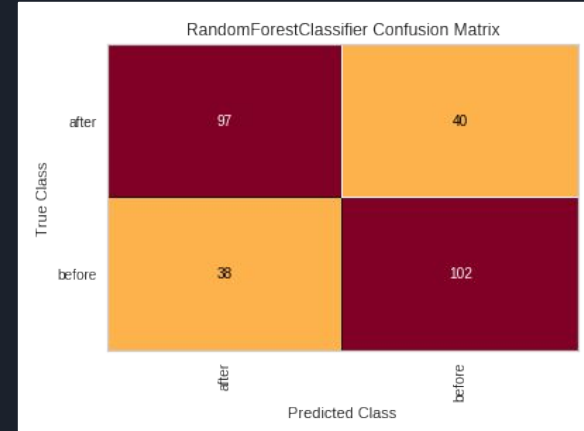
- “Before” & “After”


Data:

- Variety of topics hand labeled by Flatiron students

Models

- Random forest (best performing)
- XGBoost
- Logistic Regression
- K Nearest Neighbors





Try it for Yourself

Explore the interactive web app [here](#)



Conclusion

Next Steps:

- Gather & train on more labeled data for classifier models
- Further refine similarity rank for recommender system
- Further optimize feature extraction speed

References:

- [Graph based recommender systems](#)
- [Wikipedia article recommender system](#)