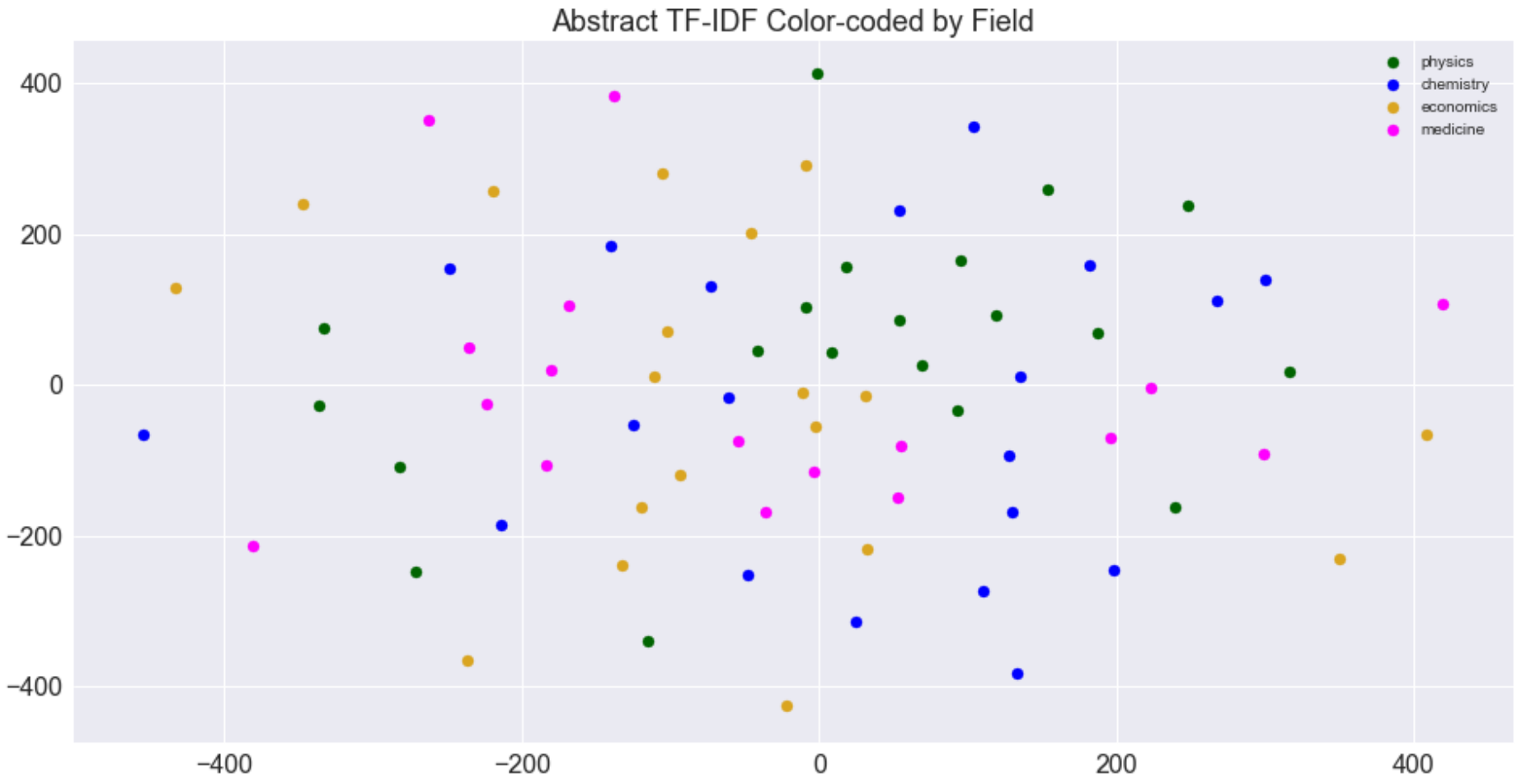
Capstone: Is A Nobel Prize Winner the Author of This Abstract?

Bronwen Cohn-Cort
The Flatiron School, Inc./Learn.co
Instructor: Jeff Herman
July 2020

Dataset

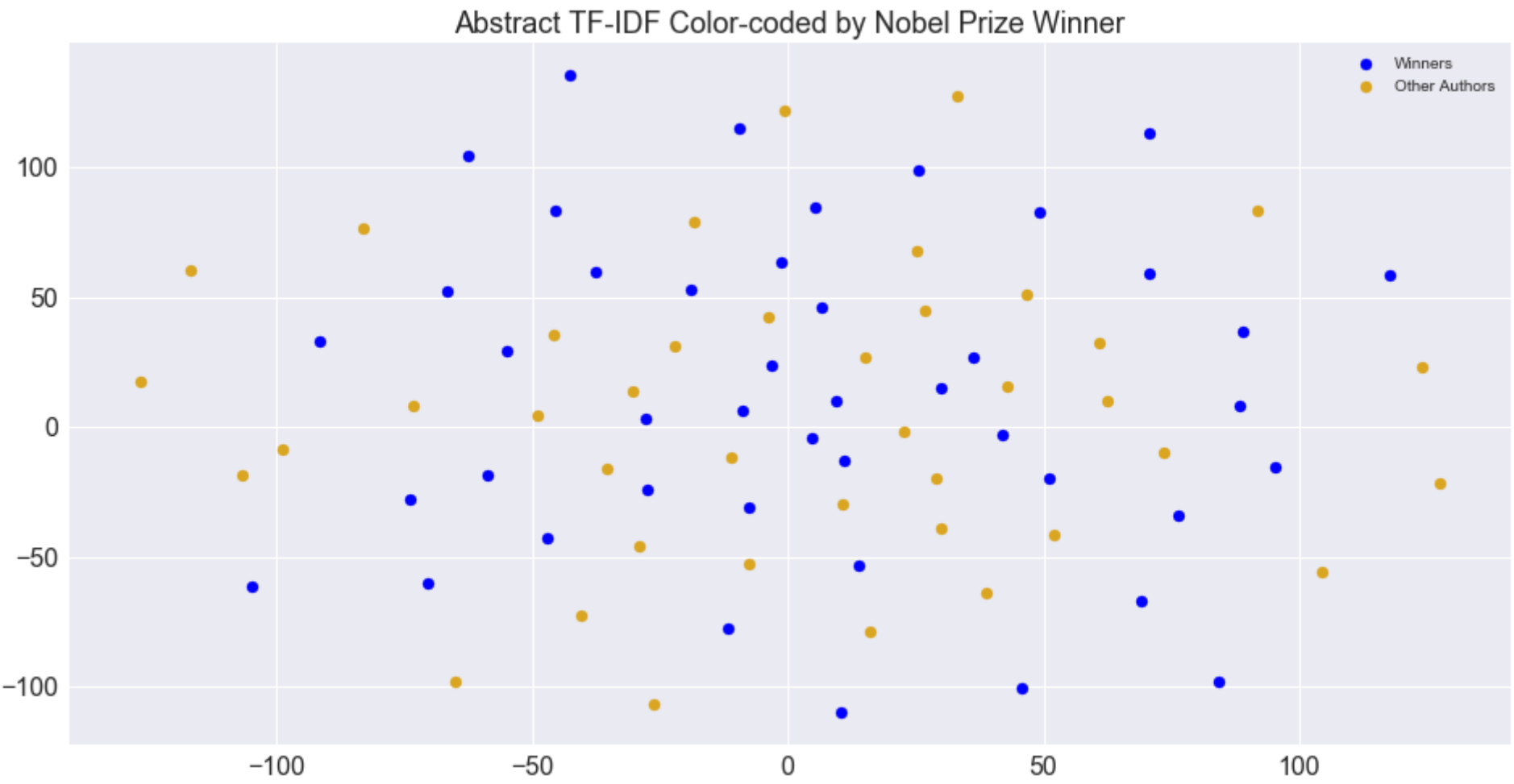
- Author information from Google Scholar
- Five most-cited publications
- Authors listed on those publications
- First five publications associated with them
- 76 records had abstracts
- Mostly even mix of Nobel Prize science fields

Abstract text color-coded by field



Abstract text by searched author

Abstract text by winner



Model Results

- Modeling by field had a low accuracy: 19% for the testing set, 30% for the training set
- Publications by field were not significantly different
- Modeling by winner had a 58% test accuracy, after improvement of the training accuracy from 38% to 58%
- There was no significant difference between publications for winners and non-winners

Takeaways

- Abstract text from different fields can be more similar than those from the same field
- Even different publications by the same author may not have abstracts that are closely related
- There are more publication abstracts for non-winners

Future Areas of Interest

- Networks of collaborators
- Text analysis for the jargon in one field
- Improved matching of author names to publications
- University/affiliation data

Thank You

Credit to the original scholarly package, which can be installed from https://pypi.org/project/scholarly/

Github: https://github.com/bronwencc/Capstone/

Many thanks to The Flatiron School, Inc. for their support

Additional Information

- Half of recent Nobel Prize winners do not have Author pages on Google Scholar, with a collection of their publications
- Abstracts may be excerpts from books instead of journal articles
- Used term frequency-inverse document frequency to analyze all abstract text
- The TF-IDF averages went into a multilayered sequential model