

Hanlin Chen
Rachel Small

Aubrey Green
Sataporn Woraissipchai

Steven Luu
Maoz Zisman

Summary

Deep Skill Stack offers multiple interfaces for users to discover and visualize online courses, allowing them to build their own course curriculum.

By combining publicly available data of online courses and creating word embedding for each course, Deep Skill Stack generated a rich similarity visualization among courses, which allows the generation of recommended courses and curriculum per user's input.

Our Approach

Algorithm

- Trained word2vec model on our dataset to generate embeddings and sentence embeddings for each course.
- Generate sentence embeddings for queries and compute the cosine similarity against our dataset to find most similar courses.
- For Course Plan:
 - Return 2 most dissimilar courses among 5 most similar courses for each difficulty (beginner, intermediate, advanced).
- For Graph: Use T-SNE dimensionality reduction to generate (x,y) coordinate pairs to visualize top courses per query.

Deployment

- Generated output datasets for each query using Google Colab and provided to PowerBI.
- Created course plan with top 2 results from each difficulty level.
- Created graph in PowerBI to plot T-SNE results and provide tooltips for each course.

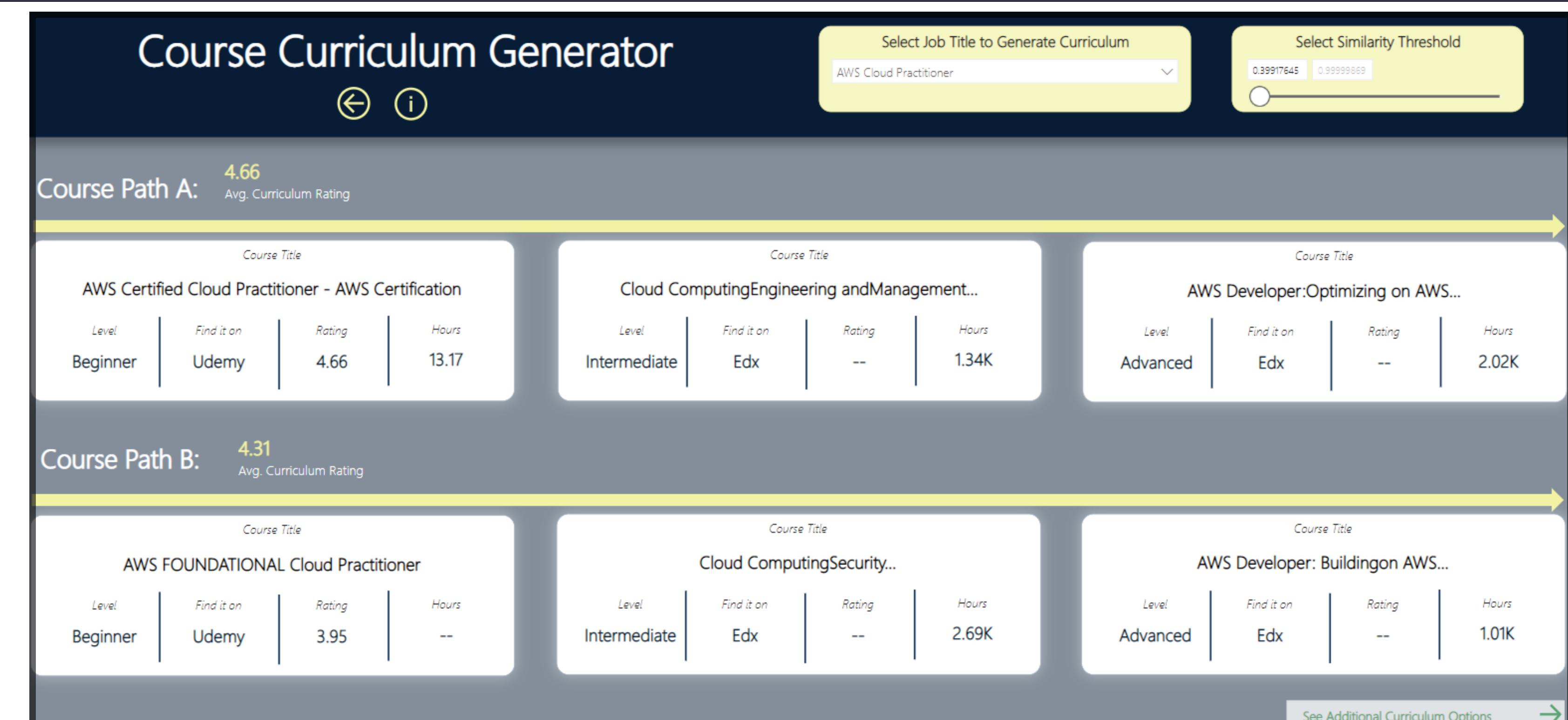


Source	Courses	Percent of Total
Udacity	267	0.20%
Skillshare	1,080	0.82%
Edx	2,365	1.80%
Coursera	3,424	2.61%
Udemy	123,915	94.55%

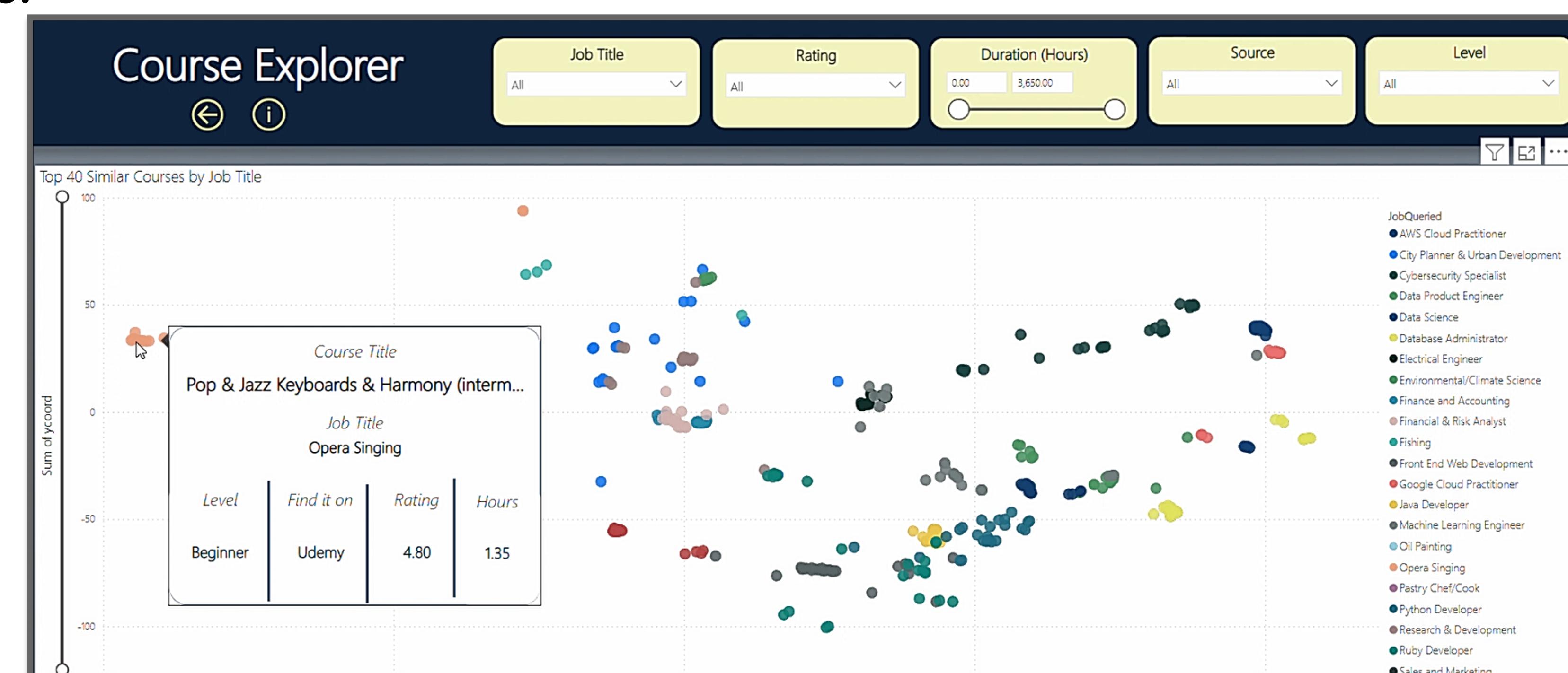
Average number of words in title & description = 23

Source: Data set was provided by https://www.kaggle.com/datasets/hossaining/udemy-courses?select=Course_info.csv, <https://www.kaggle.com/datasets/patrickgndotti/udacity-course-catalog>, <https://www.kaggle.com/datasets/siddharthm1698/coursera-course-dataset>, <https://www.kaggle.com/datasets/mahmoudahmed6/skillshare-top-1000-course>

Deep Skill Stack: A Tool for Creating an Online Curriculum using Machine Learning



- The Course Curriculum Generator returns the courses most similar to the selected job title at varying levels of difficulty.
- This allows the user to receive several options of course paths, the average rating, and the duration of each course.



- The Course Explorer Tool shows the 40 courses most similar to each job title in addition to their related proximity to one another.
- This allows for the user to browse similar courses and explore related, or non-related job titles.

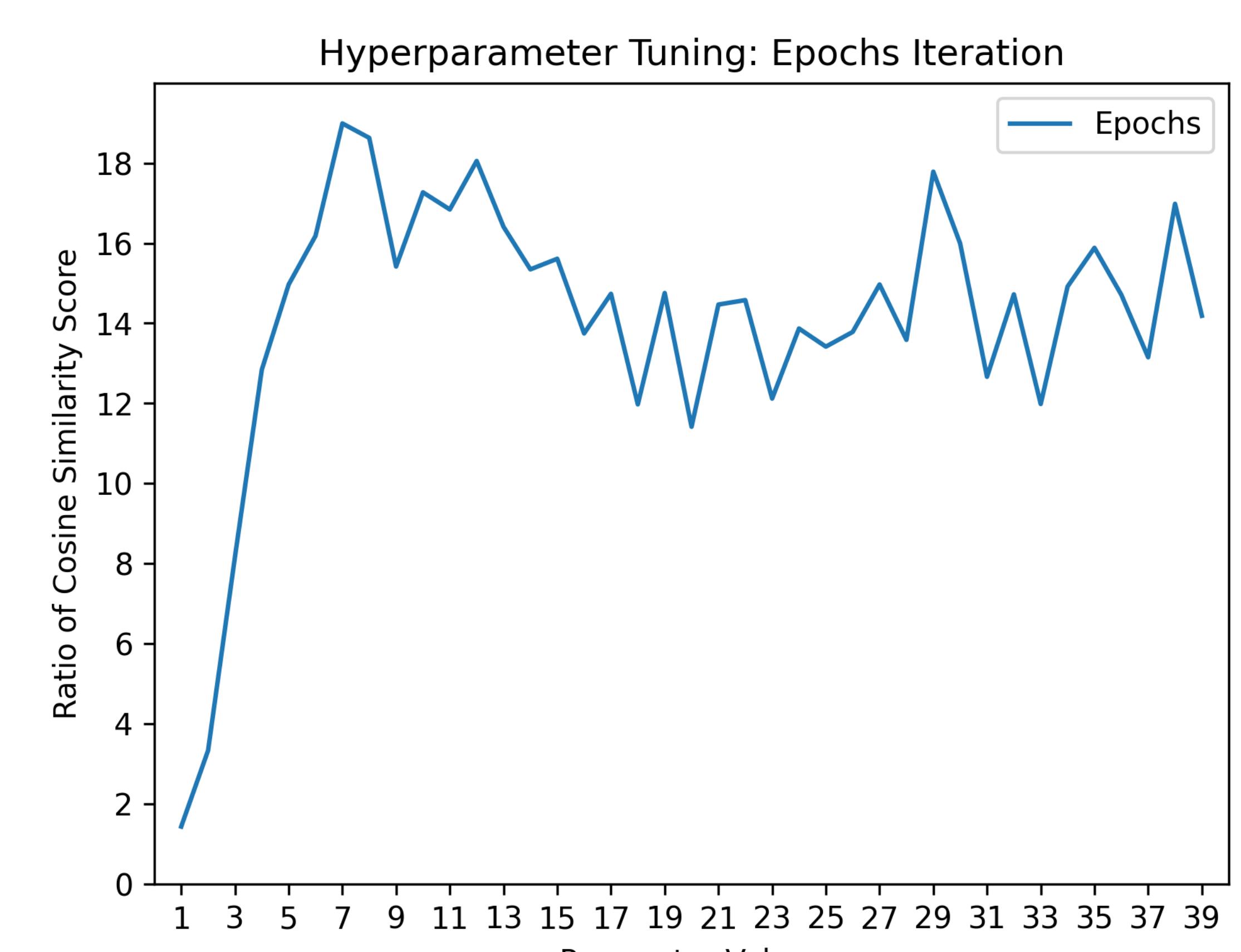
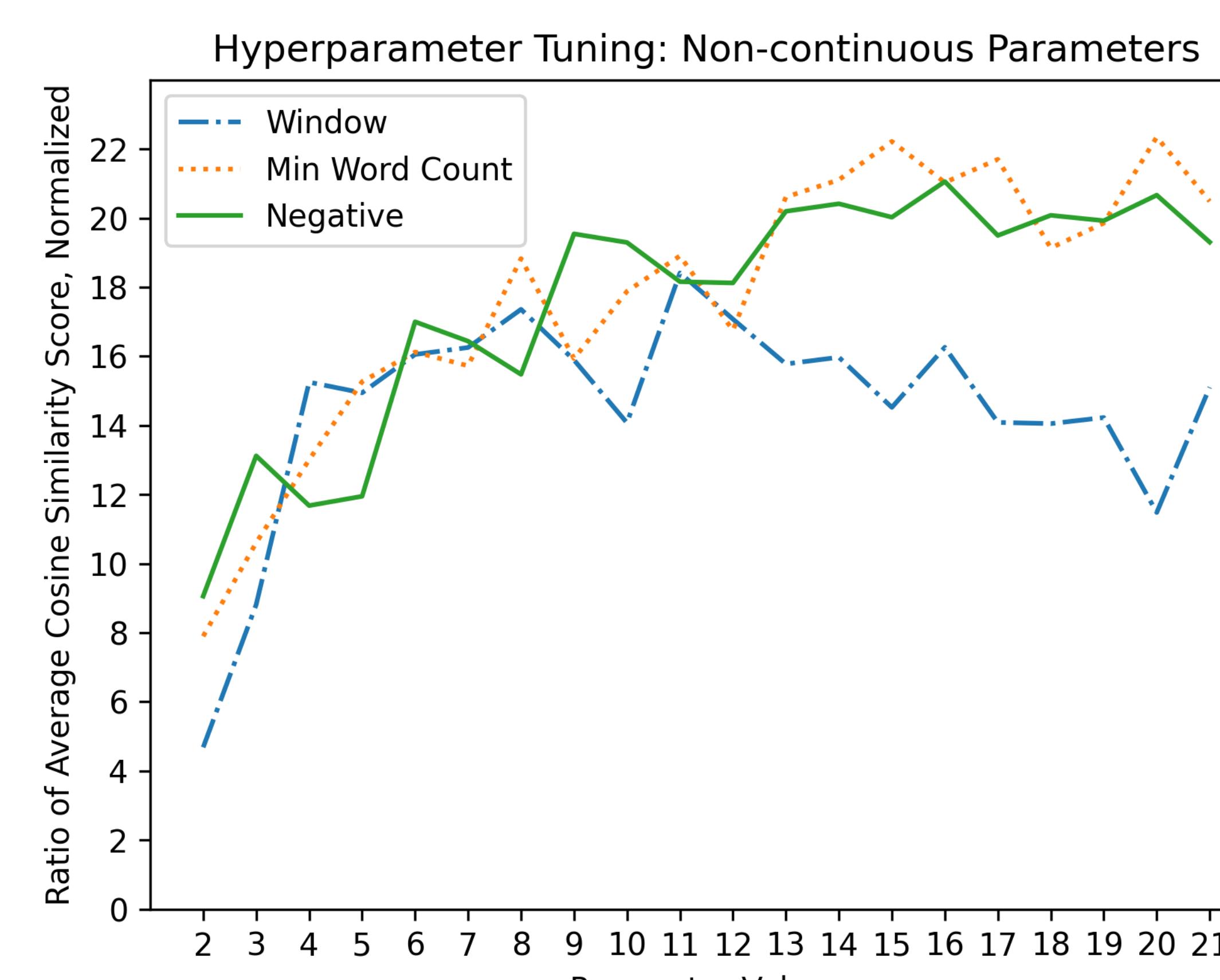
i want to learn how to swing trade cryptocurrency and stocks

Clear
Submit

output		
<code>orig_title \</code>		
42548	Penny Stocks Trading 101	
99542	Complete Algorithmic Trading Course - Forex, S...	
129660	Equity Stock Markets:Concepts, Instruments,Ris...	
124624	Reinforcement Learning for Trading Strategies	
125135	Overview of Advanced Methods of Reinforcement ...	
125794	Inflation Analysis for Investment Appraisal	
<code>orig_description level \</code>		
42548	Learn how to trade smaller stocks for bigger g...	beginner
99542	Algorithmic Trading Masterclass - Learn Forex ...	beginner
129660		NaN intermediate
124624	In the final course from the Machine Learning ...	intermediate
125135	In the last course of our specialization, Over...	advanced
125794	In this 1-hour long project-based course, you ...	advanced
<code>source duration_min rating cos_sim</code>		
42548	udemdy	295.0 4.3 0.935052
99542	udemdy	308.0 4.1 0.933982
129660	edx	60480.0 NaN 0.699520
124624	coursera	NaN 3.4 0.679561
125135	coursera	NaN 3.2 0.663541

- The web API hosted on grad.io allows the user to submit queries directly to our model to return a DataFrame or json object containing the course recommendations.

Experiments & Evaluation



The default word2vec parameter performed well by the output of the T-SNE. We try to improve the cosine similarity score by tuning the 4 hyperparameters in this experiment. Parameters that deviated from the default model (n=5) did not produce significant improvement in the cosine similarity score.