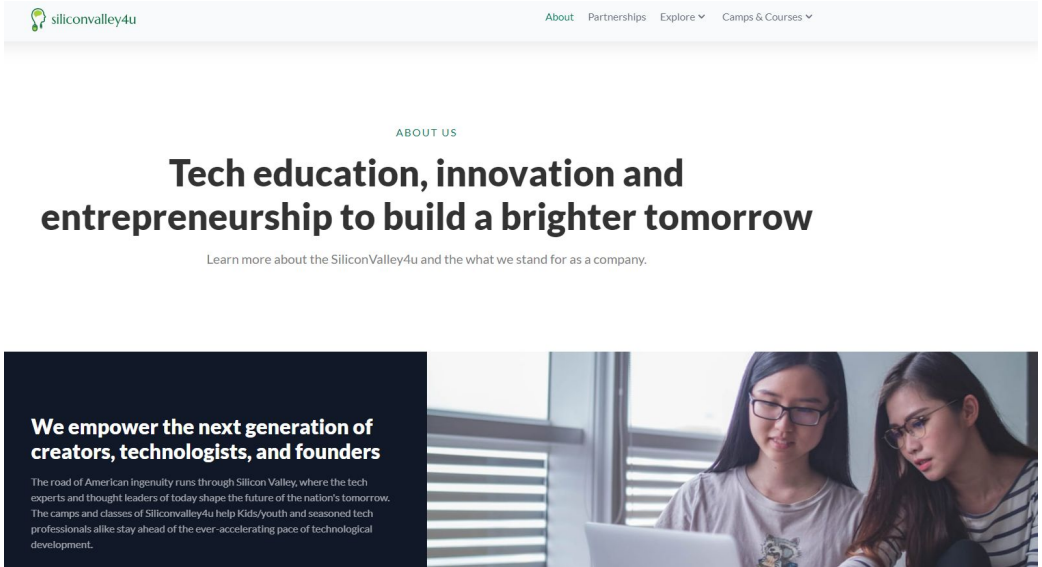
SwatCloud Recommendation System

Introduction

- SwatCloud provides educational services and resources to students who are interested in breaking into the technology industry.
- The company wants to implement a recommendation system that would match its students to an industry and specific job listings that are the best fit.



Forming the Job Database

- The first step in the project is to construct the job database itself.
- Our system automatically pulls all available job listings from ~20 top reputable companies in the technology sector.

Data Overview - Most Common Titles

- For a high level overview, we identify what the most common titles are in our job database. This allows us to form industry categories that will be one of the main outputs our model produces.

manager	2676
engineer	1768
senior	1762
associate	1093
software	936
tax	719
management	679
business	662
data	630
cloud	578
services	552
intern	547
software engineer	542
engineering	541
specialist	536
consultant	506
lead	501
program	486
2023	482
analyst	481
federal	442
sr	441
product	437
development	405
sales	398
developer	389

Data Overview - Industry Labels

- On the basis of the most common titles, we define seven distinct industry categories which are intended to cover the breadth of job types in our dataset.
- These categories are as follows:
  - Software Engineering
  - Data Analysis
  - Marketing
  - Sales
  - Operations & Finance
  - Research Scientist
  - Cybersecurity & IT

Recommendation System - Industry Matching

- The Industry Matching model assigns a compatibility score to a user based on their background.
- This model is a 4-layer neural network built with the Keras package. The layers are as follows:
  - Embedding Layer
  - Convolutional 1D Layer
  - Global Max Pooling Layer
  - Dense Layer
- These layers are all fairly standard for NLP models. Future iterations of this project are recommended to focus on the quality of the dataset over this model structure.

Recommendation System - Industry Matching

- Example test cases shown to the right:
  - Top candidate is a Software Engineer at Amazon with a background in Data Science
  - Bottom candidate is a Marketing Analytics Manager at a startup.
- This works well for younger students or students who are interested in a career change and are curious as to which industry type their experience is most compatible with.

Candidate #2: Software Engineer @ Amazon with academic background in computer science and data science

```
classify('-Worked on the backend team to develop the Edtera web application, a learning engagement platf
```

```
1/1 [==============================] - 0s 33ms/step
                        % Match
Industry
Data Analysis         55.016786
Marketing             36.207151
Software Engineering   3.333176
Sales                  3.207008
Operations & Finance   1.994777
Cybersecurity & IT     0.232973
Research Scientist     0.00813
```

Candidate #3: Senior Marketing Analytics Manager @ Rippling with extensive work history as a marketing data analyst

```
classify('1. Create measurement framework across different funnel stages (TOF, MOF, and BOF) and marketi
```

```
1/1 [==============================] - 0s 25ms/step
                        % Match
Industry
Marketing             84.602392
Software Engineering   7.485791
Data Analysis          6.054982
Cybersecurity & IT     1.055249
Sales                  0.408382
Operations & Finance   0.239769
Research Scientist     0.153403
```

Recommendation System - Direct Job Matching

- The Job Matching model directly recommends the jobs from the database which are the best fit for the student.
- This model uses Cosine Similarity and CountVectorizer - a simpler algorithm compared to the neural network used in the Industry Matching NLP model.

Recommendation System - Direct Job Matching

- The same two test cases are presented here again.
- Note that although the top candidate's work experience is focused on his current work as a software engineer, the model considers his entire academic background as a data scientist and recommends those jobs accordingly.

Candidate #2: Software Engineer @ Amazon with academic background in computer science and data science

```
top_x_recommendations(10,df,'-Worked on the backend team to develop the Edtera web application, a learning engagement
```

	Job Title	Company	Similarity
0	Federal - Senior Data Architect	Accenture	0.417321
1	Federal - Data Strategy & Management Manager	Accenture	0.380363
2	Data Engineering Solution Associate	Deloitte	0.373577
3	SAP Data Management Consultant	IBM	0.372678
4	Data Engineer Summer Intern	JnJ	0.360744
5	Senior Cloud Big Data Engineer	JPM	0.355947
6	Data Analytics Practitioner with Polygraph	Deloitte	0.355252
7	Senior Data Scientist - Machine Learning Opera...	JPM	0.351399
8	Federal - Data Strategy Senior Manager	Accenture	0.342962
9	Senior Manager - Data Engineering	HP	0.342095

Candidate #3: Senior Marketing Analytics Manager @ Rippling with extensive work history as a marketing data analyst

```
top_x_recommendations(10,df,'1. Create measurement framework across different funnel stages (TOF, MOF, and BOF) and ma
```

	Job Title	Company	Similarity
0	Marketing Manager, Advanced Compute Solutions/...	HP	0.374408
1	Performance Marketing Director	JPM	0.365424
2	Vice President - Digital Marketing Manager	JPM	0.362804
3	Vice President, Marketing - Campaign Managemen...	JPM	0.360134
4	Sr. Director, Web Marketing Campaigns	Microsoft	0.341826
5	Senior Marketing Manager, On Road Staffing - R...	Amazon	0.337890
6	Head of Chase Wealth Management Marketing, Exe...	JPM	0.337376
7	AM - SMA and Model Portfolio Marketing VP	JPM	0.335133
8	Social Media Lead - Cloud Marketing Organization	Microsoft	0.328559
9	Senior Associate, Marketing Operations	KPMG	0.326649

Improvement Areas

- A major area of improvement that should be focused on for future project versions is the method in which we select industry labels. These categories were manually defined based on human assessment of the most common words in the job titles combined with intuition on how these words are related to the most common industries in today's job marketplace.
- For these models to truly thrive, we also need to improve the efficiency with which we acquire new job listings & companies. Right now, there are ~10,000 jobs from 20 companies, but ideally we'd want this database to have 100's of companies and 100,000's of jobs.

Conclusion & Wrap-Up

- The model works as a great starting point for newcomers looking to find a role in tech (Industry Match NLP Model) and also as a recommendation model for experienced students eager to start the application process (Direct Job Matching Cosine Similarity Model)
- The system is designed to holistically evaluate each student based not only on their resume but also their full profile which includes their academic history & projects.
- This personalization fulfills SwatCloud’s business need for a recommendation system that is tailored specifically for its clientele.