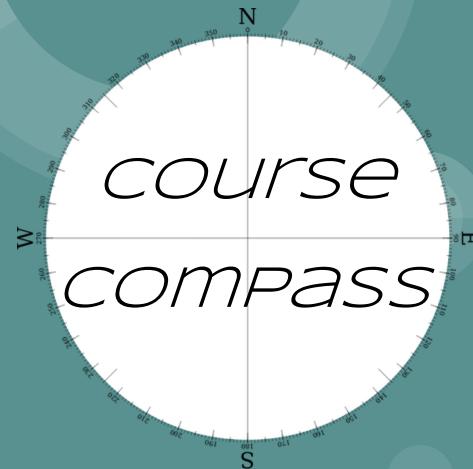


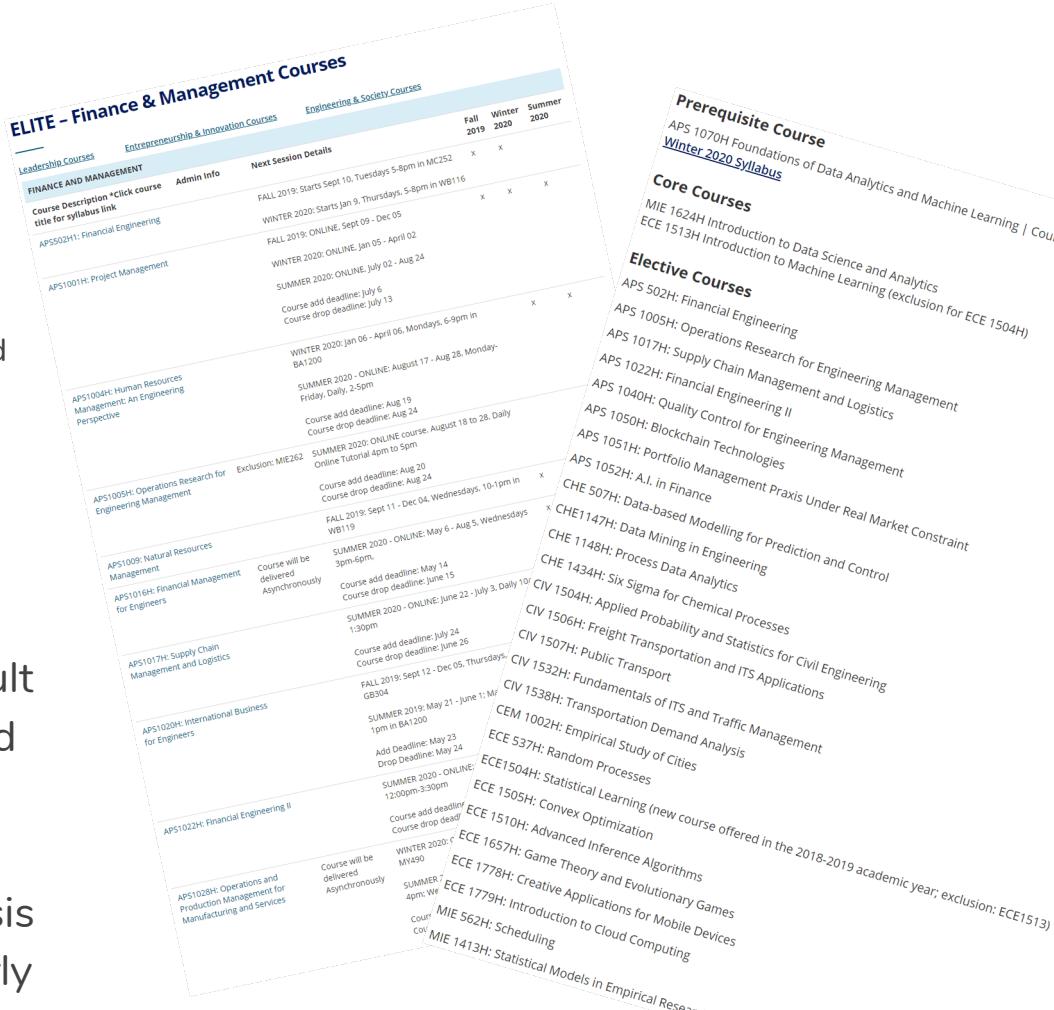
# MIE 1624 Project Presentation

Group 22: Neil Juan, Tyler Rankin, Lei Lei, Ke Ren,  
Raghavendran Thiruvengadam, Arun Shanmugam



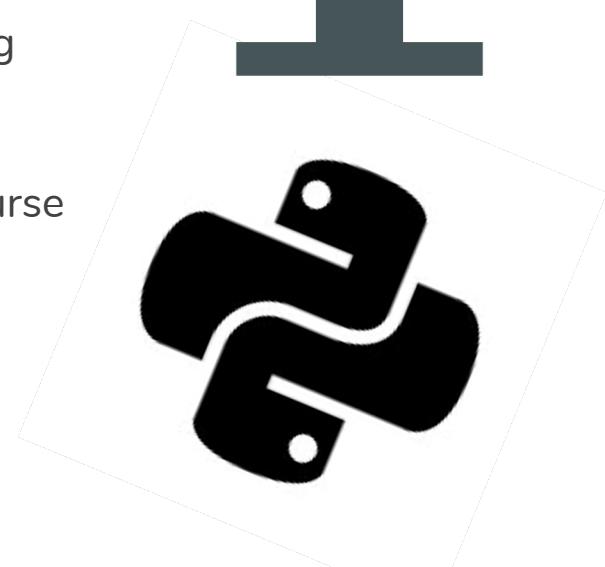
# Motivations

- Difficult for
    - Students to choose courses related to their interests and career.
    - The faculty to recommend courses to students
  - Course materials are not demanding and it is difficult to design new courses and programs.
  - The web page for emphasis is overwhelming and poorly organized.

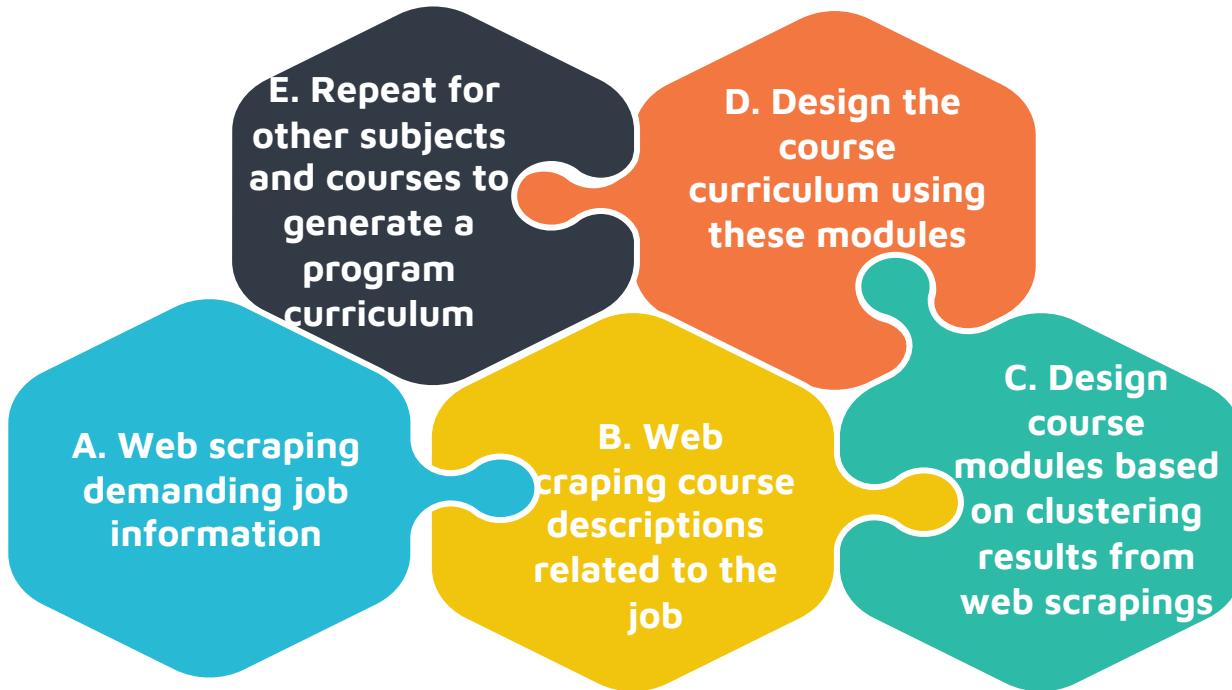


# Solution

- Design a web Python APP using machine learning and data science:
  - Help faculty to design/update course curriculum based on current job demand and other existing online courses.
  - Help students to find interesting and useful course for their future career.



# Design process



# Data Collection by Web-Scraping



## Online-Course Platforms

	Link	Description	Course Title
0	<a href="https://www.classcentral.com/course/spatial-da...">https://www.classcentral.com/course/spatial-da...</a>	Spatial (map) is considered as a core infrastr...	Spatial Data Science and Applications
1	<a href="https://www.classcentral.com/course/process-mi...">https://www.classcentral.com/course/process-mi...</a>	Learn to get a critical, process-centric persp...	Introduction to Process Mining with ProM
2	<a href="https://www.classcentral.com/course/datamanage...">https://www.classcentral.com/course/datamanage...</a>	This course presents critical concepts and pra...	Data Management for Clinical Research
3	<a href="https://www.classcentral.com/course/python-mac...">https://www.classcentral.com/course/python-mac...</a>	This course will introduce the learner to appl...	Applied Machine Learning in Python
4	<a href="https://www.classcentral.com/course/coursera-d...">https://www.classcentral.com/course/coursera-d...</a>	Dog Emotion and Cognition will introduce you t...	Dog Emotion and Cognition

- **Job Search Websites:** Indeed, Glassdoor, ZipRecruiter
- **Online Course Platforms:** Coursera, Class Central, Course Report
- **Key Words:** Data Analyst, Data Scientist/Engineer, Data Science
- **Scrapped Contents:** Job/Course Descriptions, Titles



company	job_title	link	location	salary	job_description
Buy With Impact	Data Scientist with Backend Experience	<a href="https://www.ziprecruiter.com/c/Buy-With-Impact...">https://www.ziprecruiter.com/c/Buy-With-Impact...</a>	Toronto, ON	\$85K to \$100K Annually	save\n\ndata scientist with backend experience...
Visa	Data Scientist	<a href="https://www.ziprecruiter.com/c/Visa/Job/Data-S...">https://www.ziprecruiter.com/c/Visa/Job/Data-S...</a>	Toronto, ON	Full-Time	save\n\ndata scientist\n\nvisa toronto, on\n\n...
Altus Assessments	Data Scientist	<a href="https://www.ziprecruiter.com/c/Altus-Assessmen...">https://www.ziprecruiter.com/c/Altus-Assessmen...</a>	Toronto, ON	Full-Time	save\n\ndata scientist\n\naltus assessments to...
Points International	Data Scientist	<a href="https://www.ziprecruiter.com/c/Points-Internat...">https://www.ziprecruiter.com/c/Points-Internat...</a>	Toronto, ON	Full-Time	save\n\ndata scientist\n\npoints international...

# Data Cleaning and Clustering

- Defining the **skill sets** of data science from WordCloud and Kaggle dataset.
- Searching for the defined skills in each of the courses or job posting.
- Calculating the **word frequency** of each skills in every observation.

### 3. Clustering the skills by Hierarchical Clustering using Dendrogram.

Description
Spatial (map) is considered as a core infrastr...
Learn to get a critical, process-centric persp...
This course presents critical concepts and pra...
This course will introduce the learner to appl...
Dog Emotion and Cognition will introduce you t...
...
The courses in the XSeries are designed to hel...
With the continuous generation of massive amou...
The fashion industry has a large influence on ...
This is a hands-on workshop that explores vari...
In this course you will learn the tools, techn...

### Finding the correlation between skills

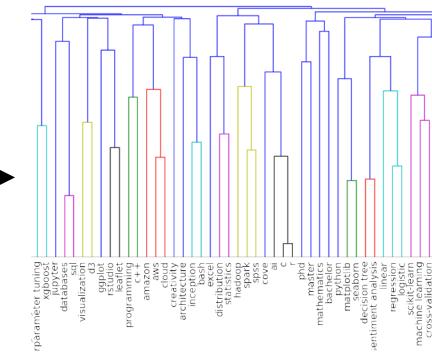
Skills-Vectorization →

	data architecture	programming	machine learning	ai	statistics	mathematic
	0.0	0.0	0.0	12.0	0.0	0.0
	0.0	0.0	0.0	17.0	0.0	0.0
	0.0	0.0	0.0	12.0	0.0	0.0
	0.0	0.0	0.0	0.0	0.0	0.0
	0.0	0.0	0.0	2.0	0.0	1.0

→ Clustering

Description Table

Word Frequency Table



Dendrogram

# Dendrogram

Clusters ➤

Skills ➤

Modules ➤

Software Engineering  
&  
Webpage Design

Soft Skills

Data Management

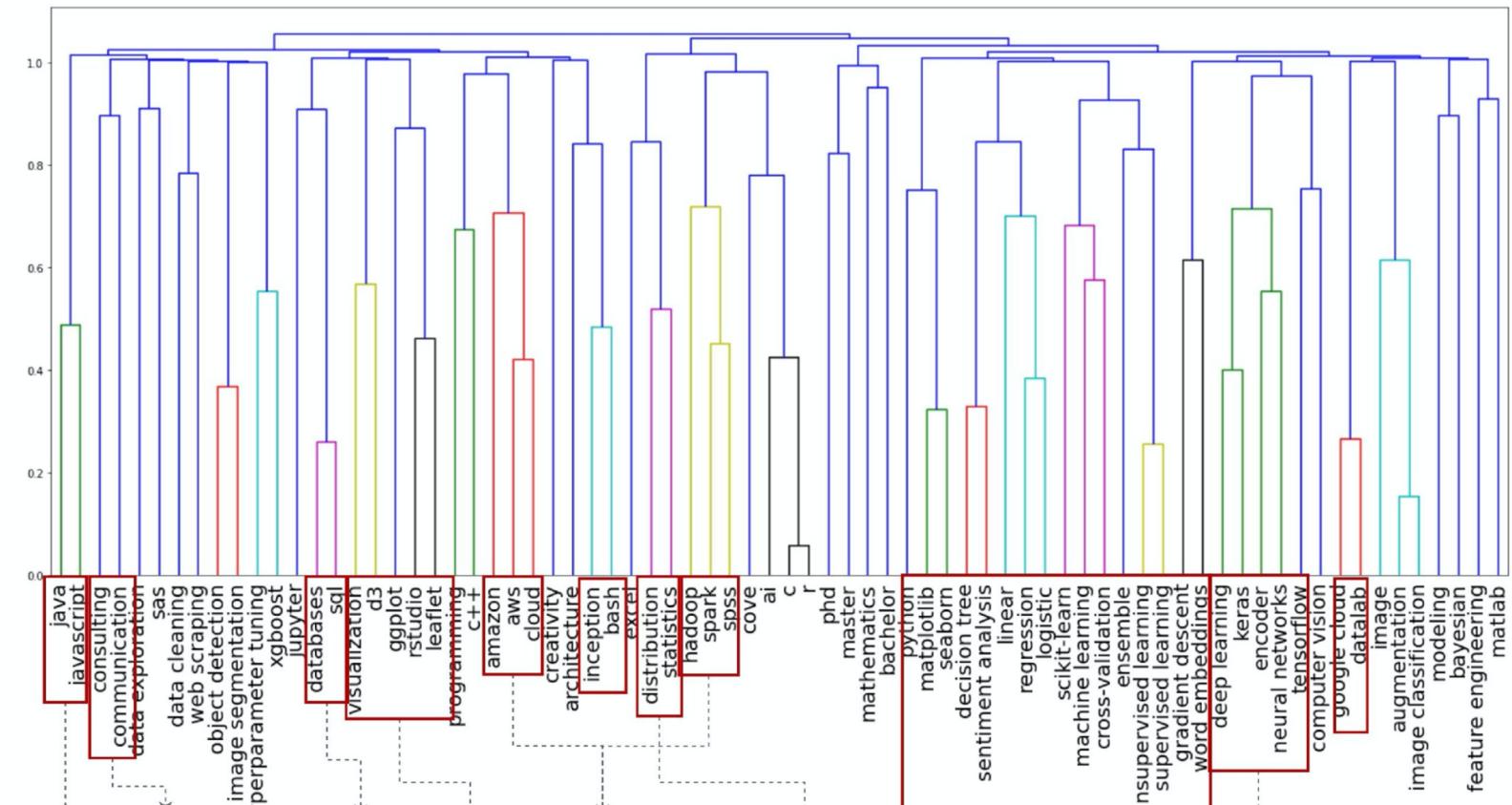
Data Visualization

Distributed &  
Cloud Computing

Statistics

Machine Learning

Artificial Neural  
Networks and Deep  
Learning

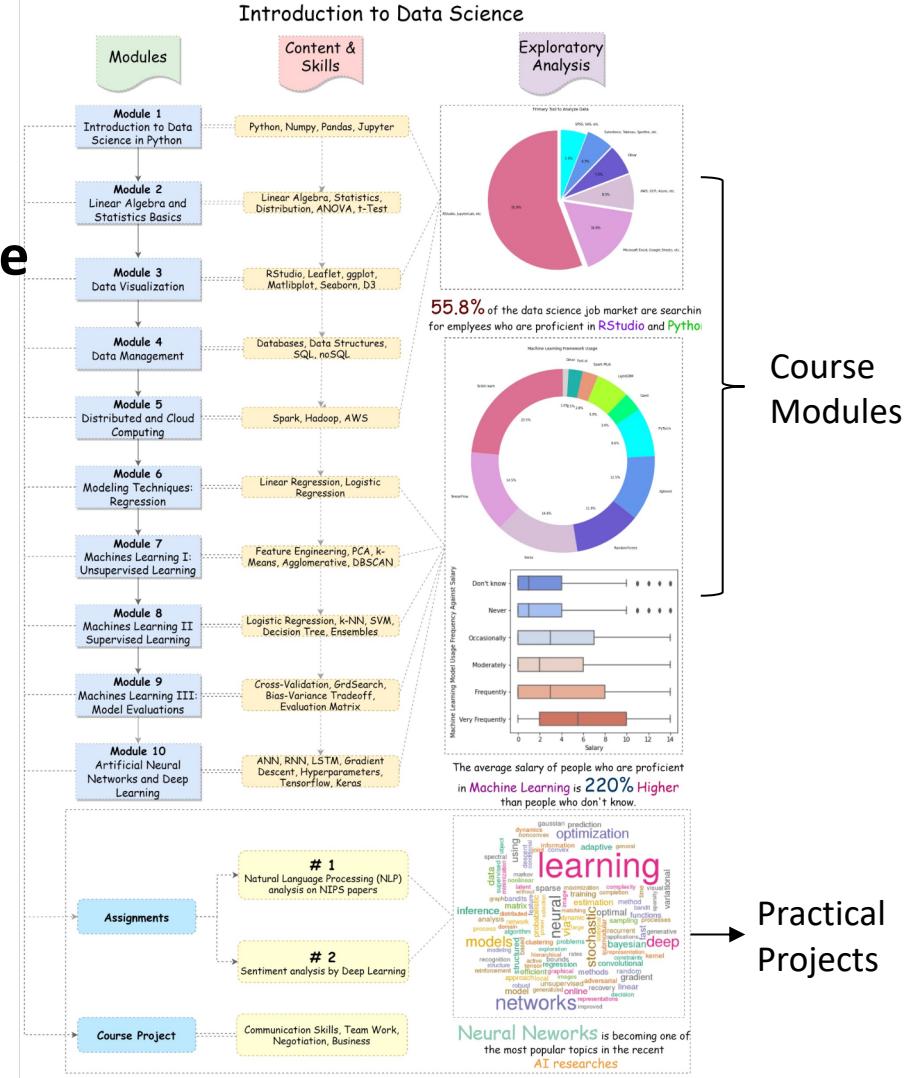


# Core Course

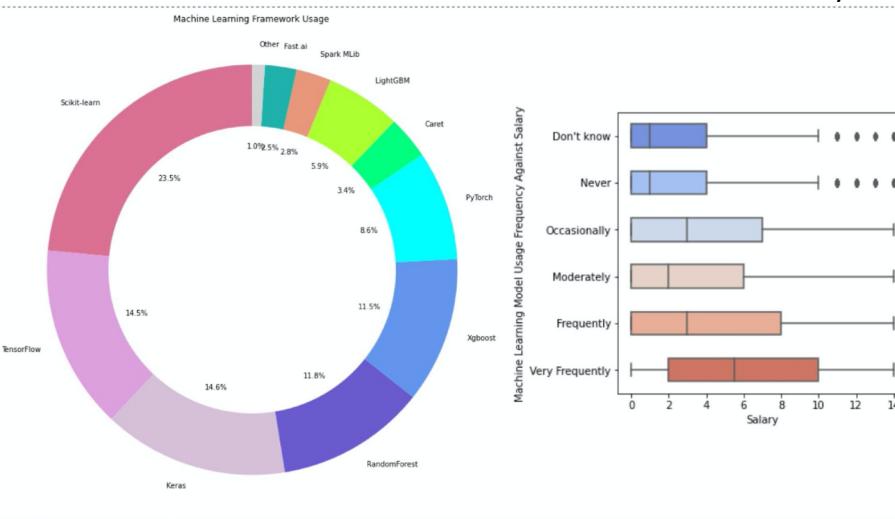
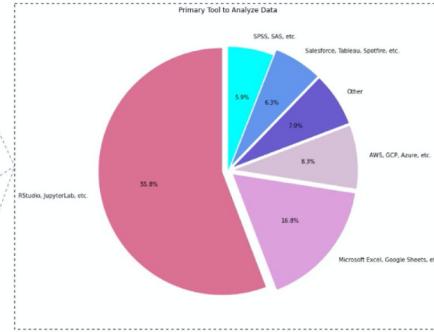
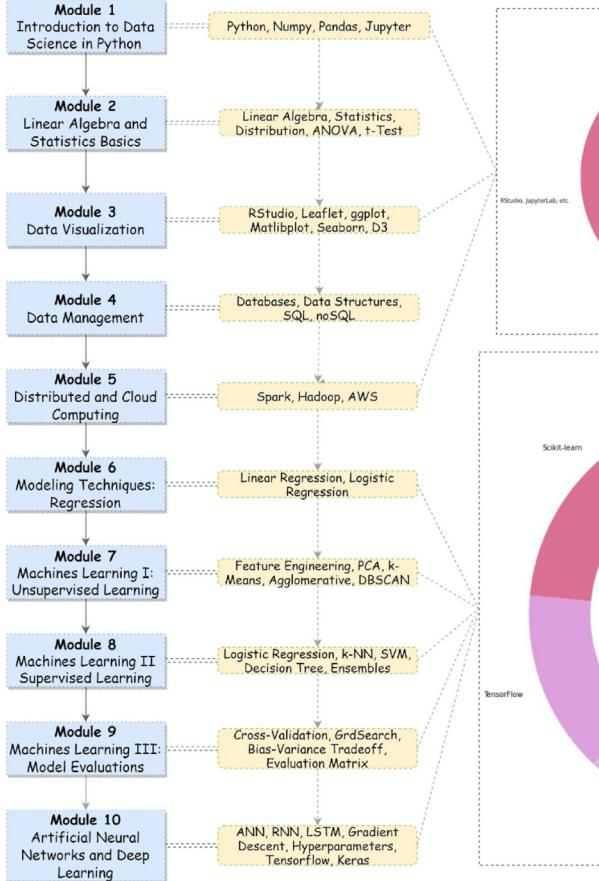
## - Introduction to Data Science

### Course Description/Purpose:

- Introduction to every possible aspects of data science.
- Emphasize the essential skills in job market.
- Provide real-word technical problem solving experience.
- Explore the applications of data science in real-world.
- Provide resources for students pursuing data science career in the future.
- Soft-skills and business sense training by accomplishing business-oriented group projects.

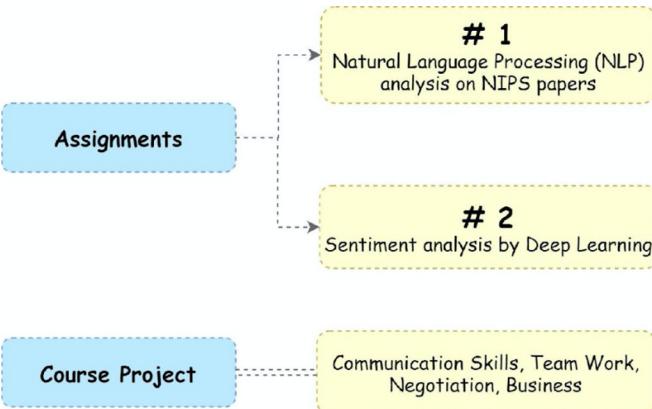


# Exploratory Analysis- Course Modules

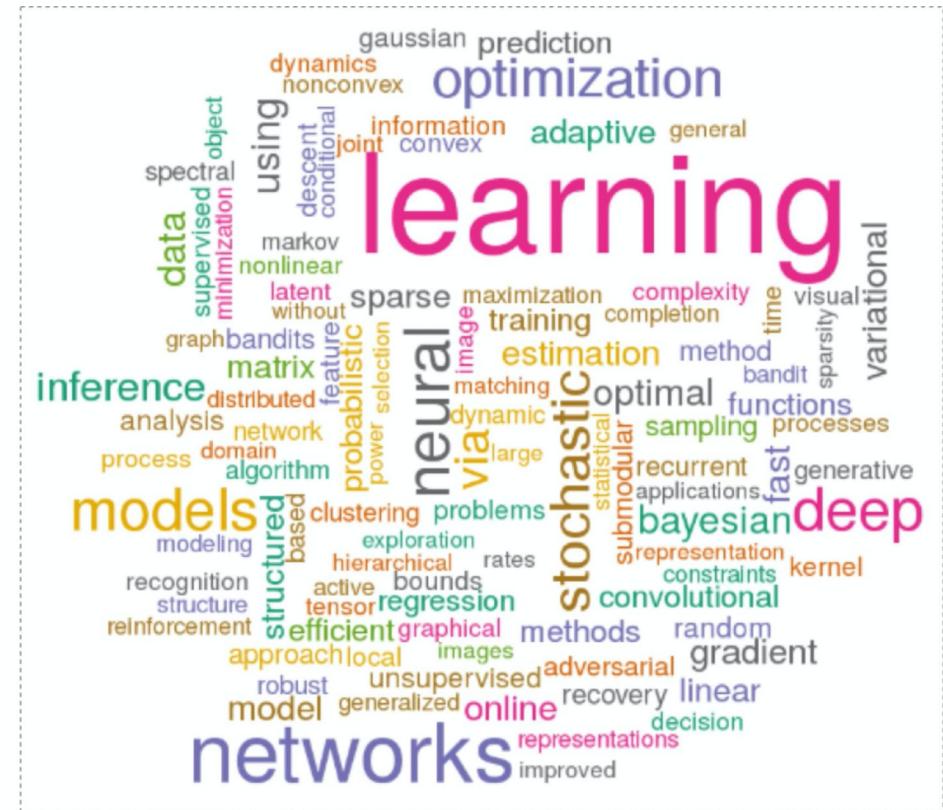


- Course Modules are selected from skill clusters.
- Python and Rstudio are the most frequently-used tool to analyze data
- Sklearn, Tensorflow and Pytorch are the most popular machine learning frameworks.
- The proficiency in machine learning is highly correlated to the salary level.

# Exploratory Analysis- Course Projects



- Assignments will focus on the most popular topics of artificial intelligence, such as NLP, sentiment analysis and deep learning.
- Group project is intended to teach students with the business-oriented soft skills.



WordCloud of the most recent research topics analyzed from NIPS

# Program Curriculum Overview



1. Introduction to Python Programming

2. Introduction to Finance

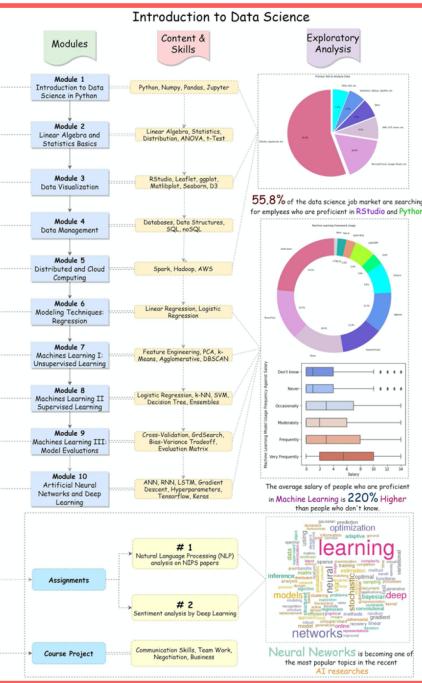
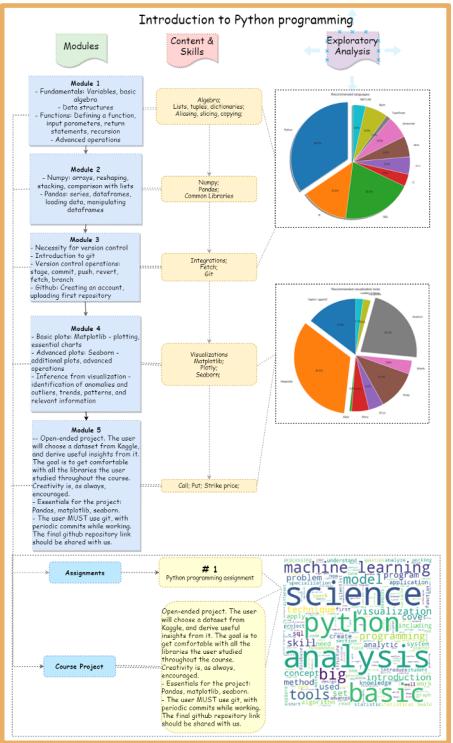
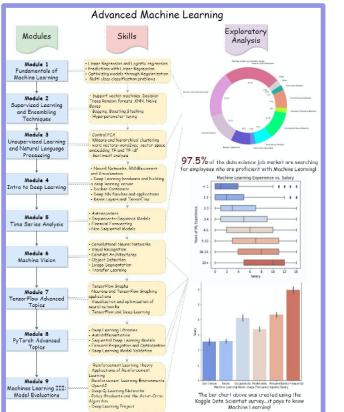
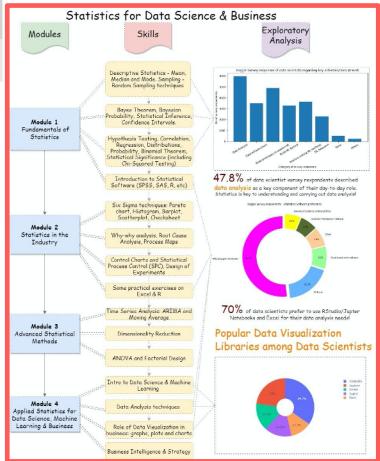
3. Introduction to Data Science

4. Introduction to Statistics and Analytics

5. Advanced Machine Learning

6. A.I. in Finance

# Course curriculum





# Business Question

- Now that the core courses and curriculum have been set, finding the correct set of complementary courses poses a new challenge
- How can we help students effortlessly choose electives?
- How can we help students using online education platforms ?

## Prerequisite Course

APS 1070H Foundations of Data Analytics and Machine Learning | Course description and time table [here](#). | [Winter 2020 Syllabus](#)

## Core Courses

MIE 1624H Introduction to Data Science and Analytics  
ECE 1513H Introduction to Machine Learning (exclusion for ECE 1504H)

## Elective Courses

APS 502H: Financial Engineering  
APS 1005H: Operations Research for Engineering Management  
APS 1017H: Supply Chain Management and Logistics  
APS 1022H: Financial Engineering II  
APS 1040H: Quality Control for Engineering Management  
APS 1050H: Blockchain Technologies  
APS 1051H: Portfolio Management Praxis Under Real Market Constraint  
APS 1052H: A.I. in Finance  
CHE 507H: Data-based Modelling for Prediction and Control  
CHE1147H: Data Mining in Engineering  
CHE 1148H: Process Data Analytics  
CHE 1434H: Six Sigma for Chemical Processes  
CIV 1504H: Applied Probability and Statistics for Civil Engineering  
CIV 1506H: Freight Transportation and ITS Applications  
CIV 1507H: Public Transport  
CIV 1532H: Fundamentals of ITS and Traffic Management  
CIV 1538H: Transportation Demand Analysis  
CEM 1002H: Empirical Study of Cities  
ECE 537H: Random Processes  
ECE1504H: Statistical Learning (new course offered in the 2018-2019 academic year; exclusion: ECE1513)  
ECE 1505H: Convex Optimization  
ECE 1510H: Advanced Inference Algorithms  
ECE 1657H: Game Theory and Evolutionary Games  
ECE 1778H: Creative Applications for Mobile Devices  
ECE 1779H: Introduction to Cloud Computing  
MIE 562H: Scheduling  
MIE 1413H: Statistical Models in Empirical Research



# SOLUTION: Course Recommender!

- Recommends electives based on the students **interests**, or **desired specialization**
- Can also suggest electives similar to a course already taken
- Personalizes the elective search by filtering courses relevant to the user key words
- Filters out previously taken courses from recommendation
- Recommends top rated online courses using dream job as input

Course Recommender

Your personal course recommender

Please enter your interests, course choices, or desired specialization

Big Data, Finance, Data Science

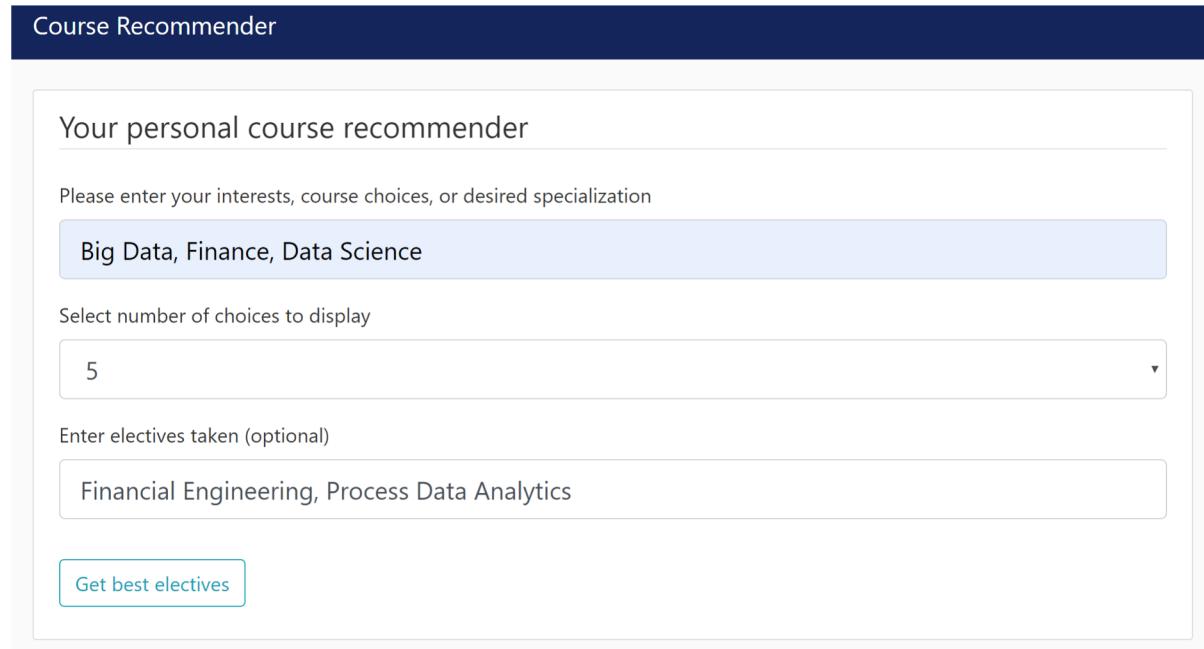
Select number of choices to display

5

Enter electives taken (optional)

Financial Engineering, Process Data Analytics

Get best electives





# SWOT Analysis

- Quick and robust application
- Pricing (free!)
- Easy to use
- Highly scalable



- System can be integrated to other school
- Additional features - weighting interests, more visuals
- Use of Deep learning for more accuracy

- Small business, inexperience
- Recommender system requires accurate course descriptions

- Required funds to scale the project
- University restrictions (?)
- Laws on scraping/information gathering