



UNIVERSITY OF TORONTO

MIE 1624 Winter 2019

Project Consulting Report

Group 12

Wenhuan Gao 1005317694

Han Jiang 1005405535

Xue Geng 1004952679

Yafan Liu 1000784948

Hyun Jun Shim 1004691942

Boqin Zhang 1000879864

Table of Contents

1. Introduction.....	1
2. Summary of Data.....	1
2.1 Job Information.....	1
2.2 Course Information.....	1
2.3 Data Cleaning.....	1
3. MIE1624 Course Curriculum Design.....	1
3.1 Lecture Design.....	2
3.2 Tutorial Design.....	2
4. Master of Data Science and Analytic Program Curriculum Design.....	3
5. Master of Management and Business in Analytics and AI Program Curriculum.....	3
Design	
5.1 Job Market and Kaggle survey Analysis.....	3
5.2 School Programs Analysis.....	4
5.3 Master of Business and Management of Analytics and AI (MBAI).....	4
6. Data Science Education EdTech.....	4
 Appendix A.....	 6
Appendix B.....	7
Appendix C.....	8
Appendix D.....	9
Appendix E.....	10

1. Introduction

Currently, the number of people who wish to learn about data science are growing ever since data science, analytics, AI, big data are becoming widely used in various fields. To ensure they are provided with vast knowledge related to the field, this project report analyzes how University of Toronto data science course and two master's program regarding data analytics or business management can be designed based on data set collection.

As students start to have interest in data science career, majority of them research what program and courses they should participate in to improve their knowledge and skills in related field. At the end of this report, it reviews the solution which area of knowledge people can focus on to build stable career in the data science.

2. Summary of Data

2.1 job information

The job information are collected from Indeed, Kaggle, LinkedIn, SimplyHired, Amazon and Monster by web scraping using keywords: data analyst, data scientist, data engineer etc.

2.2 course information

We designed course curriculums through collecting course data from:

- Coursea
- Datacamp
- Dataquest
- Queens' University MMAI program
- York Univeristy MBAN program

2.3 Data cleaning

Data cleaning includes removing noise from the data file, producing a cleaned list through removing html tags, URLs, stop words, html character codes, and transforming all texts to lowercase letters from the data set.

3. MIE1624 Course Curriculum Design

For curriculum design in Part 1, our group chose to focus on the frequency of keywords that are related to Data Science. So here we first scrapped job information and online course information from LinkedIn, Indeed, Monster, SimplyHired, Coursera, Dataquest, Datacamp and Kaggle. Frequency of keywords in course and job website are shown in Appendix A. From the top frequency words, we can know which skills will be the top rank skills that needs to be taught in Data Science Course.

We also use the linear regression model and computing the matrix correlation coefficients to train the model in order to rank the algorithms, data processing tools, programming languages and machine libraries according to their relationship rates with High salary jobs and popular courses.

3.1 Lecture Design

Lecture 1: Introduction

Here we will give a course about the history of Data Science and some methods for process big data. This lecture focus on giving general ideas and arise interests of students for this course.

Lecture 2 - 4: Introduction to Machine Learning

From lecture 2 to 4, as employers want to hire students who have strong background on Machine Learning Algorithm, we will be providing the introduction and mathematical proof of different methods and algorithms. From, the analysis in job and courses websites, we find that the most popular ML methods are Neural network, Deep Learning, Decision Tree, Logistic Regression, Clustering Algorithm, Monte Carlo Method, Random Forests, Genetic Algorithms and AI. Therefore, these five lectures will talk about these knowledge points.

Lecture 5 - 6: Introduction to Data Analysis

From lecture 5 to 6, as data analysis skills play a very important role in data science. Many employers like banks and stocks centers want to hire more students who can analyze large scale commercial data. And some technical companies such as Amazon and Google also pay more attention on the data on cloud. So here, we will give lectures to talk about data analysis in the field of Business, Web technology.

Lecture 7: Introduction to Feature Engineering

In the description of jobs in Amazon and Indeed, jobs like data scientist and data analyst require employees can analyze all kinds of features for their data like mathematical features, categorical features and image features. Lecture 7 will give students the introduction about feature engineering and popular methods used to do feature selection and feature analysis.

Lecture 8: Data Visualization

In daily works for data scientists, all of them need to know how to present analyzed data properly. This lecture will cover general visualization skills and models for data presentation.

Lecture 9 - 10: Model Implementation

For the final two lectures, as all students have already known the general knowledge about Data Science. We need to teach them how to design a good model to train our dataset to achieve our goal. Here, we will give lectures about model stimulation and model optimization.

3.2 Tutorial Design

For the design of tutorials, our group thinks it would be better to provide more practical skills to students. So, our design of tutorial will cover the practices of Python programming, Business analysis and Model Implementation. In addition, in order to help students better understand Machine Learning, our tutorials will also give expansion about the machine learning algorithm (e.g., Neural networks to Convolutional neural networks). Finally, tutorials will cover the review for assignment 1 and 2 as usual.

4. Master of Data Science and Analytic Program Curriculum Design

The job websites, such as LinkedIn and Monsters, shows the top 5 skills, which are most required for data science careers, are Big Data, SQL/databases, AI, Data Analysis, and deep learning. The coursera and Datacamp also shows that Big Data, SQL/databases, AI, and Data Analysis are searched by customers more than others. The detailed data visualization is shown in Appendix B. The mentioned subjects are the most required subjects for Data Scientist, Data Engineer, and Data Analyst. They should be part of at least core courses of the program. It also shows that some other skills are also asked as the requirements, such as neural network, data mining, visualization, and machine learning etc. Even though they are shown to be less important compared to previous four subjects, learning them could improve individuals' data skills; which is why, it would be proper to include them in elective courses.

The Master of Data Science and Analytics program is designed to be 20 months, 4 semester program for full time students. Students are required to take 10 half credit courses including summer internship with capstone project course during the last 4 months. In order to graduate, they need to finish 8 core courses with 2 electives. The detailed sequence of courses is shown in Appendix B.

For the first term, student will go over data science courses in introduction level learning analytics techniques such as regressions, simulation and optimization modeling, machine learning etc and these course will teach students about basic knowledge of Python, SQL, R language. In the second term, students will learn about introduction level of machine learning, visualization, and data mining. Machine learning will cover the basic algorithm knowledge of Regression, Classifiers, Logistic Regression, Decision Trees, Neural Networks, Clustering, etc., while visualization course and data mining will cover data selection, cleaning, and coding. In the third semester, students are provided with big data science course which will allow them to learn deeply about Big Data Architecture with an exposure to Hadoop technology. Neural Networks and Deep learning course provide in-depth knowledge of fundamental concepts and algorithms behind neural networks, including back propagation, automatic differentiation, and distributed representations etc.

5. Master of Management and Business in Analytics and AI Program Curriculum Design

5.1 Job Market and Kaggle survey Analysis:

With positions containing keywords analytics and data manager, business analyst, AI system designer searched from Indeed.ca, we observe that manager position occupies a large portion of the job market, and popular skills they are looking for from candidates are marketing, A.I. and management skills. Overall, AI, Strategy and Marketing are top 3 technical skills for positions in this field, and management, communication, and service are the top 3 soft skills that are required.

With only considering jobs that are higher than 100,000 on Kaggle survey, Data Scientist, Software Engineer and Chief Officer are the top 3 occupations with higher salaries. The best years of experience ranges from 5 to 15 years. However, for those with experience within 5 years, the more experience, the higher compensation.

5.2 School Programs Analysis:

We investigated Queens' University's MMAI program and York University's MBAN program, and collect common skills they are offering through the course. The result hierarchical cluster provides us good insight in combining skills into courses, and we take them into consideration when designing our own MBAI program.

5.3 Master of Business and Management of Analytics and AI (MBAI):

Combining the job titles and required skills from job market and surveys with existing school programs, we design our own curriculum for the MBAI program to minimize the gap between job market and education.

It is a 20-month program that spreads in 5 semesters/levels. The first level focuses on fundamental data science and business marketing knowledge development. Student also get chances to work in groups to develop their soft skills. The second level drives the students to deeper concepts in AI, organization strategies and consulting analysis. With fundamental skills that students equipped with, they will have a term for internship/co-op to explore the real industry. This is the best opportunity for them to practice and develop professional technical and soft skills, and broaden their knowledge. After they come back and knowing what specific field they are more interested in, they can choose 1 elective course that best fit their interest. Meanwhile, they will have a individual research project where they can apply AI knowledge into Finance field. Lastly, we will invite some employers from the real industry as a specialist for students' final capstone project. This provides students more chances to get a job after they graduate. Simultaneously, they keep learn more in the field they are interested in as an elective course. The detailed sequence of courses is shown in Appendix C.

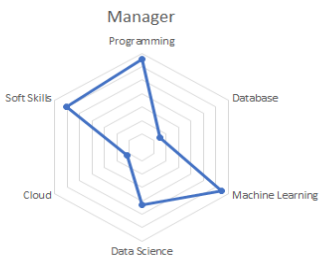
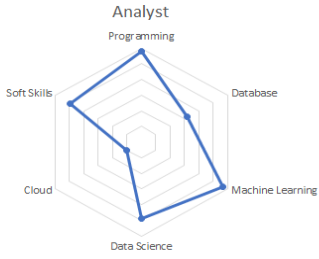
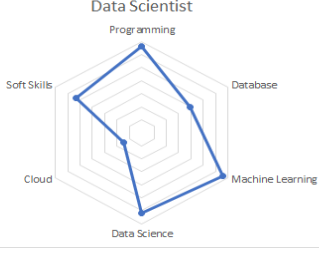
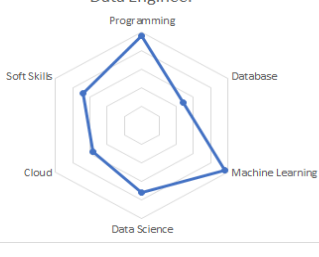
6. Data Science Education EdTech

Assuming we established EdTech startup. And the target users of our startup are:

- Master students who want to explore a career in data science area.
- People who want to improve their skills in certain data science area.

We propose the following questions with corresponding answers based on our analysis of related data:

For people who are interested in learning data science and want to develop a career in related fields, which area of knowledge they should emphasis on if they plan to become management positions, data analysts, data scientist, or data engineer.

 <p>The radar chart for the Manager role shows a high level of Programming (outermost ring), followed by Machine Learning and Database (mid-level rings). Data Science, Cloud, and Soft Skills are at lower levels (inner rings).</p>	<p>Potential candidates for manager in data-related fields should have more emphasis on programming skills, deep understanding and experienced on machine learning, as well as certain level of management skills.</p>
 <p>The radar chart for the Analyst role shows high levels of Programming and Machine Learning (outer rings). Database and Data Science are at mid-levels, while Cloud and Soft Skills are at lower levels.</p>	<p>Potential candidates for data analyst should have more emphasis on programming skills and machine learning, as well as certain level of problem-solving skills and knowledge on data science.</p>
 <p>The radar chart for the Data Scientist role shows high levels of Programming and Machine Learning (outer rings). Database and Data Science are at mid-levels, while Cloud and Soft Skills are at lower levels.</p>	<p>Potential candidates for data scientist should be skillful on programming and machine learning, as well as certain level of knowledge on data science and communication skills.</p>
 <p>The radar chart for the Data Engineer role shows high levels of Programming and Machine Learning (outer rings). Database and Data Science are at mid-levels, while Cloud and Soft Skills are at lower levels.</p>	<p>Potential candidates for data engineer should take courses that have more emphasis on programming and machine learning, as well as certain level of knowledge on data science and cloud.</p>

To further explore the relationship between job positions and types of related skills, we use K-Nearest Neighbors (KNN) algorithm on the dataset and demonstrate ranking information. Trained model provide train accuracy of 67.5% and test accuracy of 58%. (Refer to Appendix E)

Appendix A

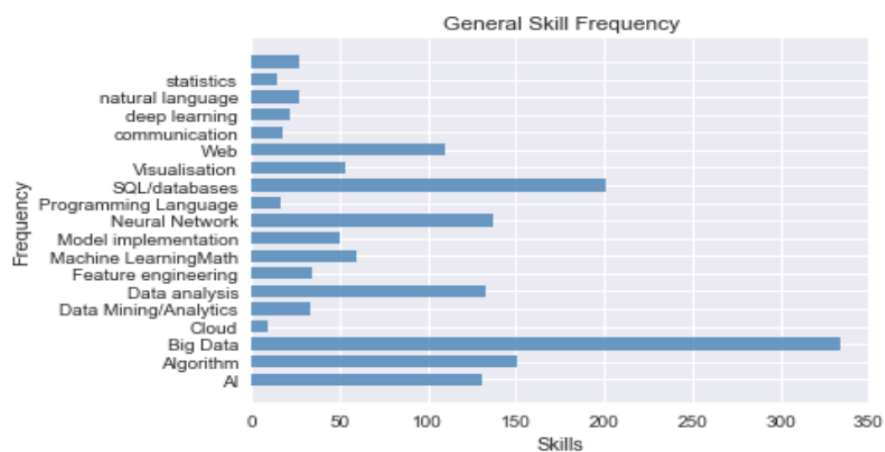


Figure 1. Frequency words from Course Websites.

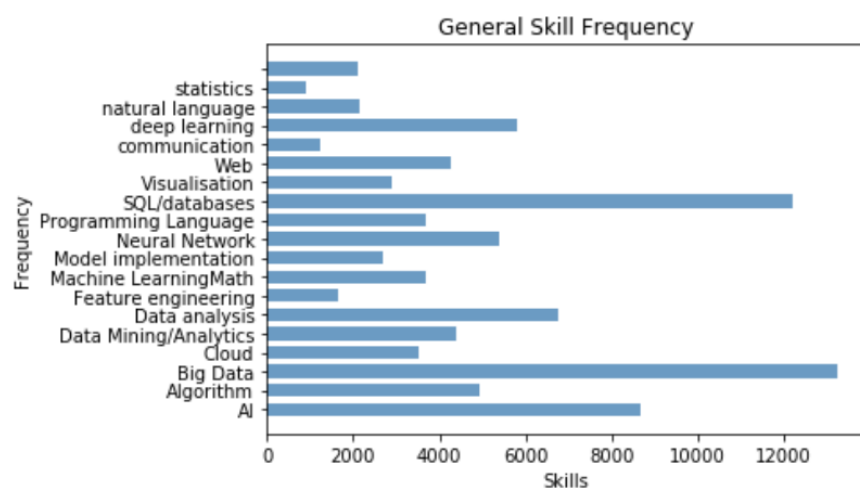


Figure 2. Frequency words from US_job Websites

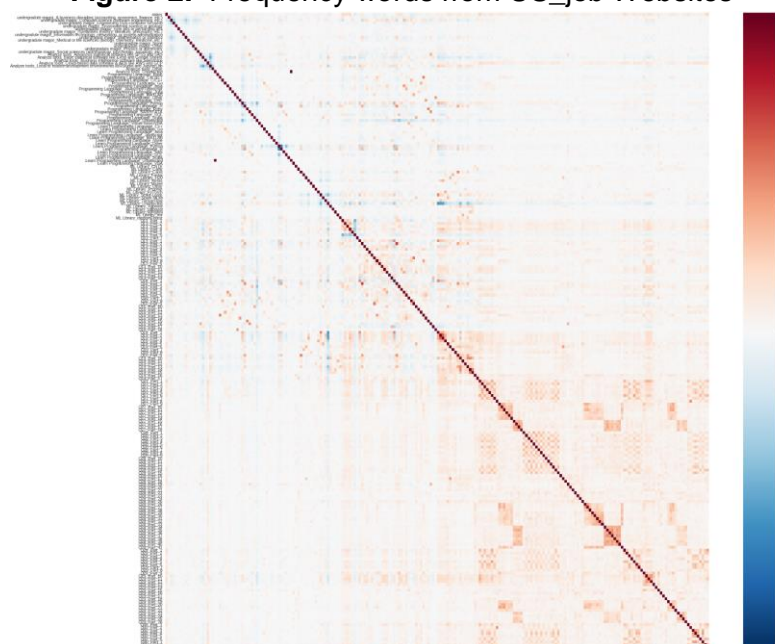


Fig3. Feature Importance for relation between skills and salary

Appendix B

Master of Data Science and Analytic Program(M.D.S.A)

Semester	Courses
1 st semester	<ol style="list-style-type: none"> 1. Introduction to Data Science 2. Probability and Statistics for Data Science 3. Introduction to Database
2 nd semester	<ol style="list-style-type: none"> 1. Introduction to Machine Learning 2. Data Visualization 3. Foundations and Applications of Data Mining
Summer Internship/Co-op	
3 rd semester	<ol style="list-style-type: none"> 1. Big Data Fundamentals 2. Neural Networks and Deep Learning 3. 1 elective course
4 th Semester	<ol style="list-style-type: none"> 1. Project Capstone 2. 1 elective course
Elective Courses	<ul style="list-style-type: none"> • Advanced Machine Learning • Neural Networks and Deep Learning • Spatial and Temporal Models • Optimization-based Data Analysis • Parallel Computing • Cloud Computing • Natural Language Computing

Appendix C

Curriculum of our MBAI program.

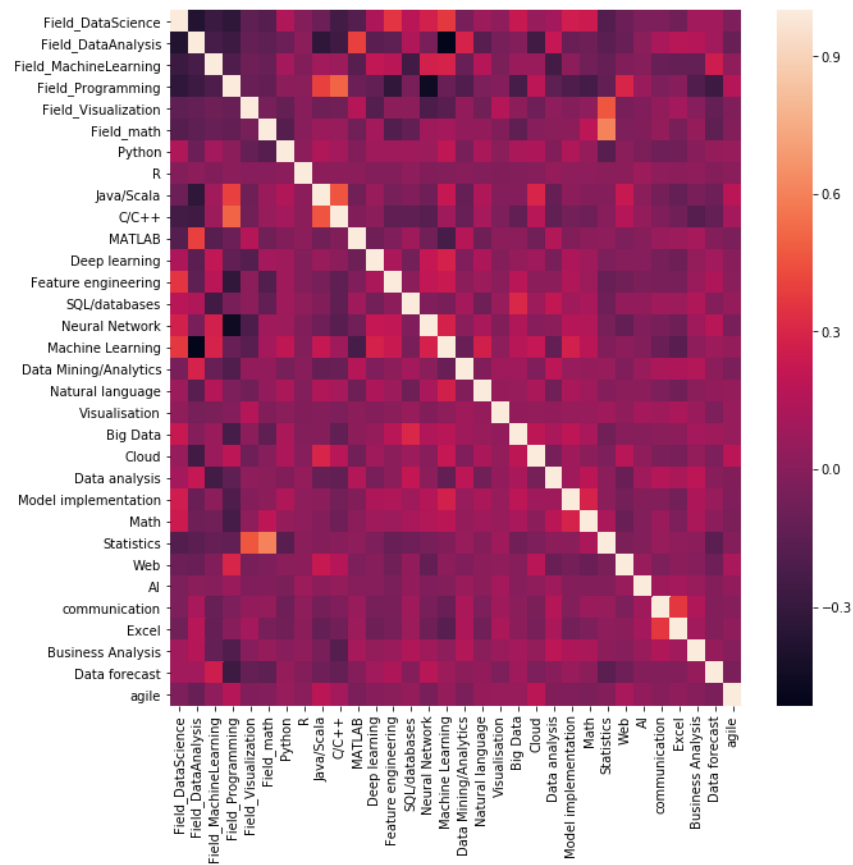
Semester	Courses
1 st semester	<ol style="list-style-type: none"> 1. Introduction to Data Science 2. Introduction to business and financial market 3. Data Analysis and Management 4. Business case study project
2 nd semester	<ol style="list-style-type: none"> 1. Deep Learning 2. Case study of an organization and market 3. Analytics consulting project
3 rd semester	<ol style="list-style-type: none"> 1. Internship/Co-op
4 th semester	<ol style="list-style-type: none"> 1. Research project to business market/AI design 2. AI in finance 3. 1 elective course
5 th semester	<ol style="list-style-type: none"> 1. Capstone with industry specialist 2. 1 elective course
Elective courses	<ul style="list-style-type: none"> • Management of Risk in Financial markets / Organization • AI in Business modeling • Optimization modeling • Consumer behavior • Case study in business study • Case study in financial marketing • Project Management

Appendix D

More figures for Part 4 observations from US market:



Appendix E



Feature Importance for relation between skills and positions.