INFO1111: Computing 1A Professionalism

2022 Semester 1

Team Project Report

Submission number: 1

Team Members:

Name	Student ID	Levels being attempted in this submission
Isaac Kim Jack Tan Ky le	510603294 ?? 510565486	Task 1, 2 ?? Level 1, 2

Contents

1. Level 1: Basic Skills
1.1. Developing industry skills

1.2. Skills: Isaac Kim : Data Science

1. Statistics

Statistics may look to first-year university students like myself to be a series of tedious calculations using fixed formulas. After all, that's what we've been doing in "Statistics" classes since high school. As a result, it's mind-boggling to grasp how statistical analysis and probability have such a profound impact on our daily lives. Statistics are used to forecast weather, restock store shelves, assess the state of the economy, and a variety of other things.(Ram Dewani, 2021) Statistics on its own is a powerful tool for getting useful insights, but when combined with computer algorithms, machine learning, or AI, it can investigate a far larger pool of data and provide answers to business, science, and societal concerns that were previously deemed unsolvable.(University of Virginia, 2020)

2. Programming Knowldege

Data science necessitates a mix of math and programming skills because it sits at the intersection of analytics and engineering. Data scientists will be unable to distinguish themselves from traditional statisticians if they are unable to use programming tools. Real data scientists will use sophisticated languages like Python or R to gain access to a wide range of useful functionality for dealing with Math, Statistics, and Scientific operations.(chandan07, 2021) Data scientists can utilise computer programmes to apply statistical expertise in a quick and secure manner, allowing them to uncover patterns in large amounts of disorganised data and make use of them.(Ram Dewani, 2021)

3. Data Wrangling

Data wrangling is the process of organising and transforming new data into the appropriate format for later analysis. It's a crucial step that's well worth the investment because it helps analysts make better data-driven judgments in less time. Outlier treatment, missing value imputation, scaling, and correcting data types can help make a data collection more orderly and aid in the development of more accurate statistical results.(Simplilearn, 2021) Data wrangling, in my opinion, is just as significant as Oxford Dictionary's decision to structure its print dictionary alphabetically. It couldn't be worse if an English learner had to go through 1000 pages of paper simply to discover a definition of one word.

4. Data Visualisation

It might be difficult to grasp what a data collection has to offer as a story when looking at a series of statistics or numerical summaries. This is where Data Visualization, the more entertaining side of data science, may be able to assist with the solution. Data visualisation enables data scientists to convey complex data sets to scholars and the general public in a more appealing and intelligible manner.(Ram Dewani, 2021) Furthermore, because computer software can automate the process, creating complex histograms or pie charts takes only a few seconds and does not require the assistance of a skilled artist.(Ram Dewani, 2021)

5. Machine Learning

Machine learning allows data scientists to easily uncover patterns in a variety of data sources and make continuous improvements to their models without the need for human involvement.(Patrick Grieve, 2022) Google and Facebook, for example, may develop highly targeted ad recommendation algorithms as more data from various people accumulates. Similarly, a skilled analyst could use data gathered by a Google search to find and predict popular software skills. On a larger scale, reliable prediction models, such as today's weather forecast system, can be constructed.(Ram Dewani, 2021)

6. Deep learning

Deep learning is a step forwards in machine learning because it uses a revolutionary method to decision-making based on a learning model that analyses data with a logical structure in much the same way that a human would.(Patrick Grieve, 2022) Deep learning applications use a layered framework of algorithms known as an artificial neural network to do this analysis.(Patrick Grieve, 2022) Using this new, sophisticated technique, data scientists may construct software for self-driving cars that can distinguish between traffic signs, detect pedestrians, and identify automobiles. (Ram Dewani, 2021) Deep learning can also be used to anticipate protein structures, create architectural plans, and design semiconductors. I am convinced that it has limitless potential for use in a variety of industries in the future.

7. Big Data Anaylsis

We produce 2.5 quintillions of data each day!(Ram Dewani, 2021) The growth of the internet, social media networks, and the internet of things (IoT) has resulted in a rapid increase in the amount of data we generate. Volume, velocity, and veracity are all high in this data.(TechTarget, 2022) If businesses do not properly analyse their acquired data in this data-rich environment, they will miss out on chances for wiser business decisions, more effective operations, bigger profitability, and happier consumers.(TechTarget, 2022) This is why data analysts are so important in today's enterprises: they help organisations minimise expenses, make quicker, more rational choices, and catch client demand at the appropriate moment to generate niche goods.(Ram Dewani, 2021)

8. Story-telling Skill

If it weren't for storytelling and visualisation, all of the data analysis and insights we create as data scientists would be pointless. Putting numbers and facts from your analysis on the table seldom gets you far. Imagine a weather forecaster walking in to tell people about an approaching blizzard. Their warning won't have any impact on the audience if they don't use appropriate visuals and storytelling techniques.(Ronald Van Loon, 2022) Hence, forecasters use graphics and interactive methods to keep viewers hooked and informed. (Ronald Van Loon, 2022)

Teamwork

I believe that the most significant element of data scientists is their ability to work in groups. Because data science is such a diverse area, it necessitates collaboration among specialists from other professions. We can't be specialists in all of the subjects and abilities listed above as individuals. It is vital to concentrate on a few abilities in which we are interested and confident while also learning from our team to improve other skills.

1.3. Skills: Tan Ky Le : Software Development

1.Programming languages

This is pretty obvious, a software developer needs to build a software from nothing and improve existing software which requires knowledge and capability over different languages. For a career as a software developer, I think one should master Java, Python and C++. Python has really simple syntax among programming languages and has been used in building mobile apps and webs, Java is a high-level language and software platform and C++ is the language for operating system programming and browsers(Mauricio Lopez, 2021). Knowing many coding languages allows more flexibility in choosing the field to work in as well as changing to different fields.

2.Debugging and testing

A software developers should not only know to code but also need to know how to test the software and fix bug. A lot of time and effort of a software project is put in testing and debugging since it is crucial that the software operates as expected. Usually, a team working on a software project has a tester to test the codes, however, one should be able to test his own code and I think it is easier to fix your code if you test it yourself. (Glassdoor Team, 2021)

3.Algorithms and data structures

It is important to understand and get familiar with different algorithms and data structures. During a software developer career, one may have to create a brand-new software or improve existing ones, thus the knowledge of algorithms and data structures will come in handy. Since the runtime and space are often required to be optimized, the developer needs to know what their software prioritize and flexibly apply different data structures and algorithms. Therefore, knowledge of algorithms and data structures will definitely be a strong skill for a software developer.

4.Version control tool (Git)

Git is extremely useful as it facilitates teamwork on a project and it allows version control which is really convenient to go back and forth between different versions of the project. Software development, which is mostly worked on as a team and consists of many procedures, can surely use git for it's convenience. Therefore, one should put time and effort in learning git, get comfortable with version control notions such as branching, merging and the command lines in git bash.

5.Teamwork skills

Software developers often work together as a team for better productivity as the work can be divided to people specializing in it. For example, a project manager who keeps things going as planned, a coder to write lines of code for the software to work. As a result, one should learn to work as a group since university. Teamwork means each person or a small group work on their allocated part of the project which offers more time, more creativity and possibilities, thus being a great benefit of teamwork. Therefore, teamwork is an important skill for people working in the industry.

6.Time management

To get the work done on time, one needs to manage and use time appropriately. It causes troubles if the work is not done on time, especially in a scenario where the work is a part of a bigger project and as a result can effect the progress of other people which is totally undesired. Good time management boosts productivity, efficiency and reduces stress while poor time managements impacts personal reputation, causes lowered motivation and energy and affect others. Thus, time management is an important skill for both life and work.

7.Self-learning ability

This is the skill we are expected to acquire from this unit. I think for any field in the IT industry, people have to do a lot of self-learning in order to catch up with the advances of technology. Technology moves on everyday and if one decides to just stand where they are, they will be left behind and replaced by others. We cannot expect schools or university to teach us all we will ever need to last long in the industry because a lot of knowledge is not yet to be available at the moment. Luckily, the internet has a lot of resources for one to learn from. Therefore, self-learning skill is one crucial skill for developers.

8.Logical thinking

This is the skill that I think is the most important skill of all. Logical thinking is the key to create efficient software codes, apply appropriate data structures and algorithms to optimize the software. I believe that it is the skill that differentiates good developers from others. Logical thinking can be acquired through learning and practices, often tiring but really rewarding for the long run.

1.4. Skills: add student 4 name here : Cyber Security

2. Level 2: Basic Technology

Level 2 focuses on initial evaluation of the tech stack that is used by a selected company. All companies make use of a range of technologies, and these technologies need to work together. A tech stack is basically just this collection of technologies that collectively enable a company's systems. As an example, one of the most common technology stacks for supporting web servers is LAMP: Linux as the underlying operating system; Apache as a web server; MySQL as the supporting database; and Perl (or more recently PHP or Python) as the programming language.

Each student should choose a different tech stack and explain the role of each of the different technologies in that stack. Note that prior to researching your proposed tech stack and spending time writing about it, it might be a good idea to check with your tutor as to whether your chosen stack is suitable. (Target = ~200-400 words per student).

2.1. Tech Stack: add student 1 name here

Your text goes here

2.2. Tech Stack: add student 2 name here

Your text goes here

2.3. Instagram Tech Stack: Tan Ky Le

1. Django

Django is an open-source web framework entirely written in Python which is great for social media sites. According to Instagram engineering team, Instagram utilizes Django and features the world's largest deployment of Django web framework(Instagram Engineering, 2016b). Django is the core of Instagram. To report errors on Instagram, the team at Disqus wrote Sentry which is a Django backed app. Django is integrated strongly helping Instagram starts quickly(Shaini A Ravi, 2020).

2. Objective-C

Objective-C can be used to create phone applications for IOS system, in this case is Instagram.

3. React Native

Instagram user app (front end) is written in React Native, a cross-functional programming language for building and developing applications on multiple platforms such as IOS, Android operating systems.(Anastasiia Lastovetska, 2021). Using React Native allows Instagram to be developed cross-platform.

4. Celery and RabbitMQ

All asynchronous activities, such as delivering notifiacations and running background processes, are handled by Celery and RabbitMQ. Celery is a task queue that is built on message communication dispersion. It's priority involves real-time activities. Celery is also helpful for scheduling. RabbitMQ is a messaging platform and is highly compatible with Celery. It is written down using the Advanced Messaging Queuing Protocol (AMQP)(Shaini A Ravi, 2020).

5. PostgreSQL and Cassandra

Instagram mainly uses PostgreSQL and Cassandra as back-end systems. Both have mature replication frameworks that fit as a globally consistent data store(Instagram Engineering, 2016a). PostgreSQL is used to contain Instagram's data such as images, videos, articles and user information(Shaini A Ravi, 2020). "Most of our data(users, photo meta, tags, etc..) lives in PostgreSQL" said the engineering team. Instagram uses Cassandra as key-value storage, assisting user photo feed, direct messages and fraud detection as well(Joab Jackson, 2018)

2.4. Tech Stack: add student 4 name here

Your text goes here

3. Level 3: Advanced Skills

Level 3 focuses on more advanced technical skills (LaTeX and Git) and analysis of linkages and relationships between the items in the company tech stack.

The following is a list of advanced Git and LaTeX skills/features. Each student should select one pair of items from each list and demonstrate actual use of each item (either through activity in Git, or through including items in this report). (Target = ~100 words per student for each feature).

- Git

  - Rebasing and Ignoring files
  - Forking and Special files
  - Resetting and Tags
  - Reverting and Automated merges
  - Hooks and Tags

- LaTeX

  - Cross-referencing and Custom commands
  - Footnotes/margin notes and creating new environments
  - Floating figures and editing style sheets
  - Graphics and advanced mathematical equations
  - Macros and hyperlinks

3.1. Advanced features: add student 1 name here

Explain your use of the advanced Git and LaTeX features.

3.2. Advanced features: add student 2 name here

Explain your use of the advanced Git and LaTeX features.

3.3. Advanced features: add student 3 name here

Explain your use of the advanced Git and LaTeX features.

3.4. Advanced features: add student 4 name here

Explain your use of the advanced Git and LaTeX features.

4. Level 4: Advanced Knowledge

Level 4 focuses on analysing your particular tech stack and considering alternatives. Each student should consider the tech stack they described for Level 2, and then discuss each of the following points:

- What are the strengths and limitations of this stack? (Target = ∼200 words).

- What alternatives exist, and under what situations might these alternatives be a better choice? (Target = ∼200 words).

4.1. Advanced Knowledge: add student 1 name here

Your text goes here

4.2. Advanced Knowledge: add student 2 name here

Your text goes here

4.3. Advanced Knowledge: add student 3 name here

Your text goes here

4.4. Advanced Knowledge: add student 4 name here

Your text goes here

References

Anastasiia Lastovetska. (2021). *APA how to make an app like instagram: Tech stack, features, costs.* (See `https://mlsdev.com/blog/70-how-to-develop-an-app-like-instagram`)

chandan07. (2021). *APA python for data science (6th edition).* (See `https://www.geeksforgeeks.org/python-for-data-science/`)

Glassdoor Team. (2021). *APA top software engineer skills to impress employers.* (See `https://www.glassdoor.com/blog/guide/software-engineer-skills`)

Instagram Engineering. (2016a). *APA instagration pt. 2: Scaling our infrastructure to multiple data centers.* (See `https://instagram-engineering.com/instagration-pt-2-scaling-our-infrastructure-to-multiple-data-centers-5745cbad7834`)

Instagram Engineering. (2016b). *APA web service efficiency at instagram with python.* (See `https://instagram-engineering.com/web-service-efficiency-at-instagram-with-python-4976d078e366`)

Joab Jackson. (2018). *APA instagram supercharges cassandra with a pluggable rocksdb storage engine.* (See `https://thenewstack.io/instagram-supercharges-cassandra-pluggable-rocksdb-storage-engine`)

Mauricio Lopez. (2021). *APA 6 reasons c++ is still in use today.* (See `https://www.jobsity.com/blog/6-reasons-c-is-still-in-use-today`)

Patrick Grieve. (2022). *APA deep learning vs. machine learning: What's the difference? (6th edition).* (See `https://www.zendesk.com/au/blog/machine-learning-and-deep-learning/`)

Ram Dewani. (2021). *APA 14 must-have skills to become a data scientist (6th edition).* (See `https://www.analyticsvidhya.com/blog/2020/11/14-must-have-skills-to-become-a-data-scientist-with-resources/`)

Ronald Van Loon. (2022). *APA the value of mastering data visualization and storytelling for data scientists (6th edition).* (See `https://www.simplilearn.com/value-of-mastering-data-visualization-storytelling-for-data-scientists-article`)

Shaini A Ravi. (2020). *APA instagram - tech stack.* (See `https://medium.com/shaini4020/instagram-tech-stack-f19ddd4dcc0d`)

Simplilearn. (2021). *APA what is data wrangling? benefits, tools, examples and skills (6th edition).* (See `https://www.simplilearn.com/data-wrangling-article`)

TechTarget. (2022). *APA big data analytics (6th edition).* (See `https://www.techtarget.com/searchbusinessanalytics/definition/big-data-analytics`)

University of Virginia. (2020). *APA how much do data scientists need to know about statistics? (6th edition).* (See

https://datascience.virginia.edu/news/how-much-do-data-scientists-need-know-about-statistics)