CASE STUDIES IN DATA SCIENCE (COSC2669)

Fortnightly Task 3

SUBMISSION DATE : 21 September 2020

STUDENT NAME : Rashbir Singh Kohli

STUDENT ID : s3810585

Option A: Interview

Name of the person I interviewed: Avinash Gaur
Who they work for: Acidaes Solutions Private Limited(CRMNEXT)
Their role: Machine Learning Lead
Linkedin profile: https://www.linkedin.com/in/avinashgaur/
Date and Time: 19/09/2020, Saturday at 04:00 PM (AEST) or 19/09/2020, Saturday at 11:30 AM (IST)
Approximate length of the interview: 25 Minutes

The questions you asked and their responses

Q   What is data science?

   Ans. He was working in the industry for a very long time and he told me that term "Data Science" is somewhat a newly coined term while itself field is as old as 50-60 years. He told me that people are working with data even way before the advancement in python by simply using excel and producing analysis graphs and helping the organisation. For him, data science is just a term that is coined to define a different subfield in computer science that uses the programming to make a statistical inference from the available data and programmatically implementing statistics.

Q   What is the best programming language for data science?

   Ans. There is no language dependency for data science, each language has its pro and con. He has used all type of languages in this field from C++, MATLAB to JAVA and Python. There are still researchers that use MATLAB for computer vision technology, JAVA for building pipelines for stream data processing(Apache Beam)(As I worked in CRMNEXT as a Data Analyst before my masters where we used Apache beam and python, so Avinash gave the reference of that), and nowadays python is famous because of its ease of implementation and libraries that make data science pipeline easy to implement. Also, I was told to not to be language-dependent as I am fresh to data science field and freshers usually do the mistake of being a language-dependent data scientist.

Q   How can a person become a data scientist?

   Ans. He gave this answer with a great laugh. He told me that he is in this field for 8-9 years now and he still can not say that he is a full-fledged data scientist. Data science is a growing field with new research or methods that are being developed daily and one can not just stop learning. Getting a masters degree in data science does not guarantee that the person is qualified or is fully fledge data scientist.

Q   Is getting an international masters degree worth it for data science field?

   Ans. No, as he already mentioned that even after 8 years he is still learning so no masters or even PHD is not the end. Data science is the field of continuous learning and the international degree can be proven useful if the user wants to enter in that countries market and also for many MNC's masters is the minimum requirement for a fresher in data science field so, if you look from that point of view then it is worth getting an internal masters degree.

Q   People are researching neural network and high-end algorithms to solver various problems while in our masters we are still just learning classification and other basic algorithms than how can we compete in the market?

Ans. In industry, people are still just using basic classification and clustering techniques but researchers are using neural networks that are not scalable or production-ready due to:

- High graphics and CPU consuming tasks.
- Are not production-ready.
- Are not verified and proven its success in various situations.
- Skill is not available to develop and maintain it in production.

So he told me that there will always be a requirement for fresh skills as the field is still growing and every fresh brain brings fresh ideas and he called data science as the field of ideas.

Q   What are the suggestions you would give to a fresher in the data field?

Ans. "Maths and statistics", He answered. Learn the maths and statistics behind the algorithms, spend as much time as you can to understand the plain maths behind the data analysis tasks. He said, "Understand maths behind the algorithm so well that you can visualise the data moving mathematically and graphically before you start coding or using libraries like Sklearn". Yes, you can brute force and hyper-tune the parameters but when you will tell the project owner about why you selected those parameters then saying hyperparameter tuning to every task you perform do not give much credibility. To learn and study that you can even tell your parents what you are doing.

Q   How should a fresher start learning?

Ans. Data science is like physics, there is a huge difference between conceptual and theoretical knowledge and practical so, "Do projects and work on real data as much you can, participate in hackathons and learn by working". You will learn an algorithm with like 6-7 features but in the real world you can have like 1000 or more features and using brute force or theoretical knowledge will not work in every case. Also, he suggested reading research papers in free time as they will help freshers to learn from other's mistake.

Q   What he expects from the person while taking an interview for his team?

Ans. To know what they are doing, did and why. Have done some project no matter even if they used only KNN in that project. But they should be able to take him through from scratch to final result and be able to reason why and what he did. A person should be confident in having a discussion and give reasonable arguments. He said he do not look for an ace of all trades but, "jack of all trades, and ace of one".

It was a good call session, as I worked closely under Avinash as a Data analyst at CRMNEXT where we were working on developing a product called DataNext(https://datanext.io) for a year and he guided me and grilled and polished my data skills and also taught me what it is like to be a professional data scientist and that being a professional data scientist does not mean that we are only going to perform machine learning throughout our profession, with his help I also learned other skills like DevOps(Docker, Rancher, AWS and server setup), worked on automated testing and developing in a scrum board based environment. He suggested that I should go to Australia to pursue my master's degree in data science at RMIT University and it was great talking to him after a long time.