# Program Assessment

# **Vinay Nagaraj DSC680 T302**

# The topic Data Science is such as vast domain and I had heard a lot about it in the past decade. I work in the IT division of a banking domain company and get to work around huge amounts of data on an everyday basis. I have also noticed the credit card industry is leaning towards machine learning based fraud detection tool from traditional rules-based tool.

I completed my bachelor’s degree in 2009 and to start school after 10 years is tough. I was searching for the right kind of course that would work in my available time and came across the Data Science MS program that Bellevue University offered. It met all my necessary criteria and hence I found this is best suited for my back to school.

My primary objective from this course was to gain a basic understanding of data modeling & problem solving with the data. At the end of this course, I can say I have a good understanding of how we can approach solving a problem using available data.

**Prepare datasets for solving problems.**

I first came across the importance of Data preparation in DSC-500 when we got to understand the concept of CRISP-DM methodology. During DSC-540 I got to work on data from multiple sources like API, web scraping and database. This course was very well structured, and the assignments were broken down in parallel with the progress of our study material. That really helped me work on my assignment and gave ample time to study and understand what code I am writing. All courses post DSC-540 has helped me to get better at looking at data and cleaning them as per my project needs.

**Recommend appropriate data modeling techniques to test hypotheses.**

I learnt the concepts of Data Modeling from DSC-550. I got to work on a few projects in every semester DSC-550, DSC-630 and DSC-650 in which I learnt and understood different modeling techniques. At the end of the course now, I’m able to differentiate between classification problem, regression problem, time series etc. Also, while doing the 3 term projects, I got a chance to apply different data modeling techniques on my datasets. Now if we get a problem to solve with data science, I will be able to identify where it will fall i.e., whether it’s a classification problem or regression problem and able to apply algorithms correctly.

**Communicate data science results into answers for domain challenges.**

Visualization and storytelling are the two main ways to convey the data science findings to general audience and executives. DSC-640 consisted completely about how we can create visualizations and tell our story to our audience. This is by far one of most enjoyable and fun classes I have had in this program. Assignments on Airline safety project was designed keeping in mind the different type of audiences and each submission used to challenge me to think differently on how I should structure my story and visualizations.

**Identify ethical considerations in dataset preparation and modeling.**

In all our courses, its clearly described what level of public data we can use. We have been taught how to generalize the survey responses, so that it doesn’t contain sensitive information. We were introduced to several techniques like PCA transformation to mask the sensitive information after preserving the underlaying relations. One of my projects was the Credit card fraud detection which had some sensitive data, and the dataset was PCA transformed when it was uploaded to Kaggle.

**Where could the program have been stronger?**

This program has made me confident when it comes to topics around data science. A couple of things I felt which would have appreciated a lot would have been more communication between everyone included in the class. This would have provided a more classroom feeling, and many concepts would have been much faster understood. I felt probably a weekly online class where the professor could have explained some of the important concepts of the week and the students to discuss the progress would have been very beneficial.

I highly recommend the change from blackboard to teams for our discussion posts as it helped students to communicate with each other and help in problem solving.

**Are there any gaps or content you wish had been covered and wasn’t?**

I feel there could have been more focus and dedicated weeks for assignments on tuning the ML models and the parameters involved in tuning. How each parameter could impact the outcome. That is one area which I struggled a little bit and had to google a lot.

**Are there any suggestions for improving the program?**

It would be very nice to have VM access probably from DSC-550 or DSC-630 and after. By this it will help us to choose big real-world datasets and work on those. There are a few limitations on our personal laptops and having VM access would have helped me a lot.

**What are your next plans after completing the program?**

I am working with my manager currently to find me a suitable role in my company for my skills. It will be tough for me financially to start as a fresher in another company. I am working on ways I can integrate my knowledge gained into my current job since it involves dealing with huge amounts of data.

**Do you feel the program prepared you for being a data scientist?**

Definitely. This course has laid a good foundation and I am confident I can build upon this knowledge gained. I am comfortable around the concepts of data gathering, data cleaning, visualization, model building activities.