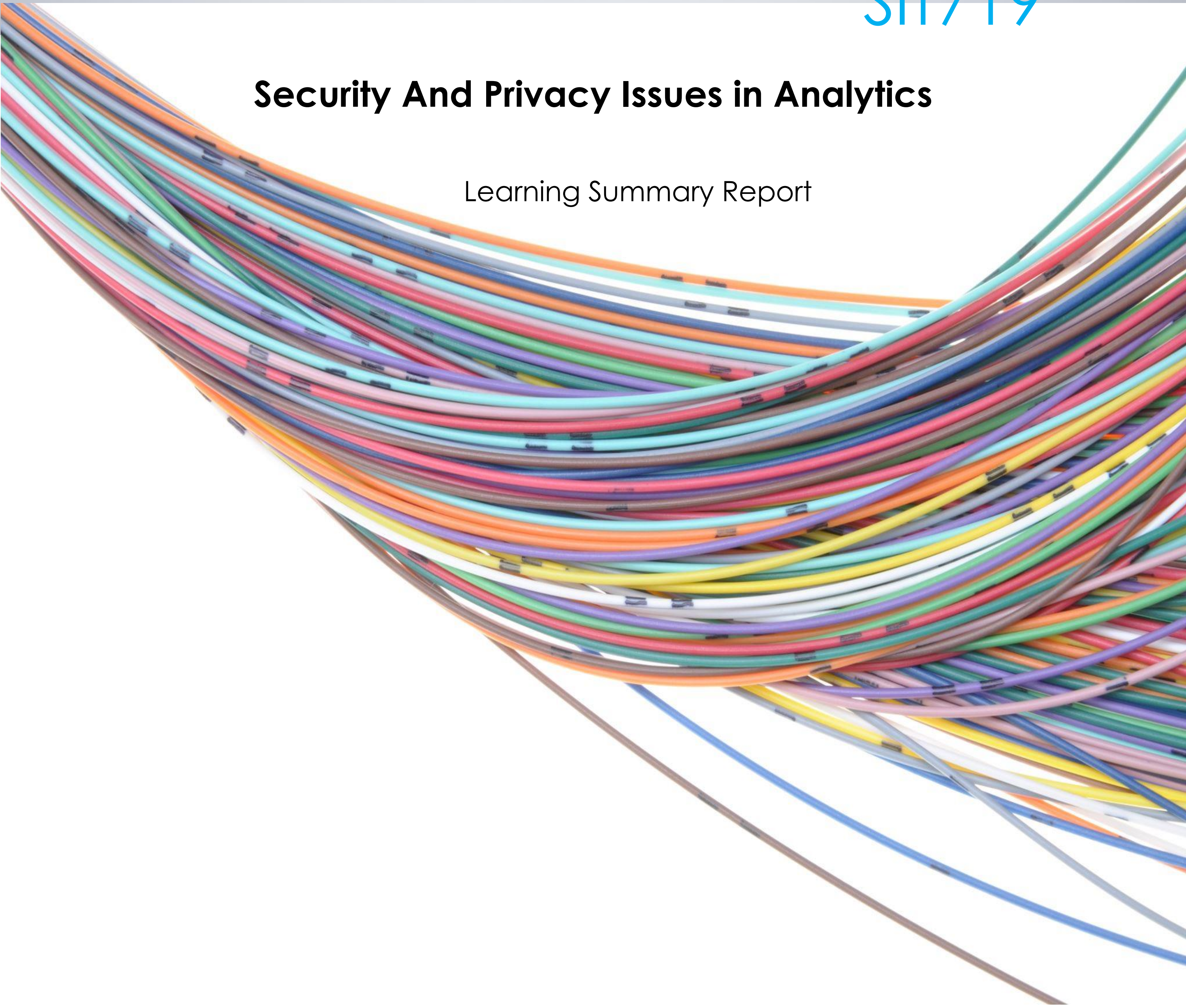
SIT719

Security And Privacy Issues in Analytics

Learning Summary Report



Prateek Singh

221218743

SELF-ASSESSMENT DETAILS

The following checklists provide an overview of my self-assessment for this unit.

	Pass (D)	Credit (C)	Distinction (B)	High Distinction (A)
Self-Assessment				✓

SELF-ASSESSMENT STATEMENT

DECLARATION

I declare that this portfolio is my individual work. I have not copied from any other student's work or from any other source except where due acknowledgment is made explicitly in the text, nor has any part of this submission been written for me by another person.

Signature: **Prateek Singh**

PORTFOLIO OVERVIEW

This portfolio contains all the work that demonstrates that I have achieved all the Unit Learning Outcomes for, SIT719 – Analytics for Security and Privacy, to minimum Pass level and aiming for a High Distinction Level.

I started learning this unit while having a Machine Learning Engineer Background. This unit taught theories that are directly applicable to the software I write. I learned about the Security and Privacy issues which affect the deployment of a Machine Learning Model in the real world. The skills learned during this unit have not only enabled me to know more about problems and state-of-the-art solutions concerning Security and Privacy issues but also allowed me to apply these skills to real-world datasets. The Distinction and High Distinction tasks also allowed me to freely explore the cutting-edge work being done in this domain. This has allowed me to employ these skills in the coursework as well as in my workplace.

The starting tasks (*1.1P – 3.1P)* were fundamental in developing hands-on scripting skills for using machine learning and other related libraries. I also learned how Machine Learning can be effectively used for designing models concerning system safety *(via Task 4.1P Attack Classification using Naïve Bayes Algorithm* and *Task 4.2C Intrusion Detection using Supervised Learning Techniques)*. The use of the WEKA tool for quick benchmarking was also a key skill acquired.

In *Task 5.1D End-to-end project delivery on cyber-security data analytics,* I employed various supervised learning algorithms and explored how supervised learning can be used for classification-related tasks for protecting computer networks. This task also taught me model evaluation and choosing the best model based on evaluation metrics for such problem statement. This task also provided an opportunity to replicate the environment and results claimed in a survey paper on a real-world dataset, this skill is very critical in developing baseline solutions and choosing appropriate models for any problem statement.

Task 8.1P Test your privacy knowledge and *Task 9.1 D/HD: Location-based Privacy Protection* introduced me to concepts of data privacy and allowed me to explore state-of-the-art measures employed in Machine Learning and Deep Learning algorithms to ensure data privacy. A key insight was the concept of *differential privacy* which taught me how to tackle data privacy issues with the right set of tools and the shortcomings of traditional methods. These tasks also allowed me to look deeper into the developments across business and legislative domains that are driving the consensus towards robust frameworks required for data protection and privacy.

In *Task 10.1P ACS Report on Privacy Preserving Data Sharing Frameworks* I explored the *Five Safes data analytics framework* which sets the premise for the environment required for developing artificially intelligent algorithm which operates on sensitive data. This task also allowed me to learn about the risk associated and the tradeoff between data privacy and data sharing. This also reinforces the observation that there is a growing global agreement to establish firm measures to ensure data privacy.

During these tasks, I have displayed hands-on capability as well as strong research temperament to look for methods beyond the coursework by conducting an independent literature survey. I believe this makes me a suitable candidate to achieve High Distinction.

REFLECTIONS

Data privacy is a relatively new frontier, with the advent of cutting-edge Artificial Intelligence algorithms unforeseen threats to individual privacy are occurring, in form of data leaks and hacks. The expectation from this unit was awareness about these scenarios, and real-world applications of their countermeasures. This unit has provided me a comprehensive understanding of privacy risks and state of the art defenses employed against them. The *Ontrack* tasks were fundamental in providing opportunity to work on real world datasets as well as explore novel works.

After completing all the *OnTrack* tasks I am very confident in my hands-on ability to code up a baseline solution for problem statements associated with data privacy and machine learning. I am also confident in my ability to comprehend state of the art methods of this domain.

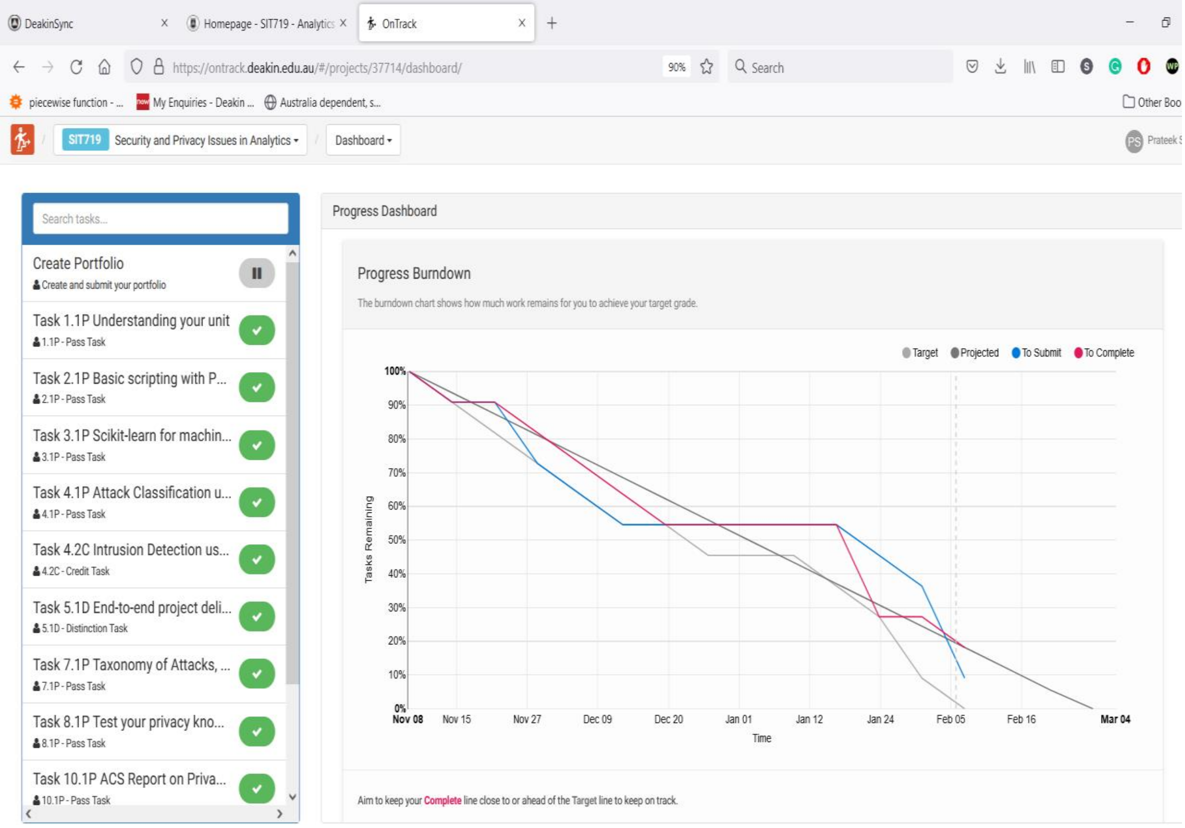
The most challenging topic for me was understanding *Differential Privacy*. This required additional efforts and reading up material beyond the coursework to understand underlying concepts. Mastering this topic taught me patience and the relevance of literature surveys. After completing the HD task associated with this topic, I am very confident about my hold over this topic,

Differential Privacy and *Task 7.1P: Taxonomy of Attacks, Defenses, and Consequences in Adversarial Machine Learning* was the most interesting topics of this unit for me. I learned risks and countermeasures associated with data privacy. This was a new concept for me which taught me how various models that are deployed in real-world are prone to such attacks.

Along with the technical skills that I acquired during this unit, one of the most important skills which I picked up was effective time management. This taught me the importance of estimating the time required to come up with solutions for a problem statement. This is also a skill which I would like to improve in the future.

The material provided on the Cloud Deakin portal and the weekly lecture helped a lot in picking up key ideas. These were complemented by the weekly workshop sessions which helped me to put these skills into practice. A major highlight was the *Ontrack* tasks, especially D and HD tasks, which allowed me to explore the literature beyond the scope of coursework and allowed me to develop pipelines on real world data.

The *OnTrack* screenshot denotes that I have taken up all level tasks and tried to finish them to the best of my ability. It also shows that I could have finished some of the tasks sooner which would have allowed me time to revisit them.



Create Portfolio
Create and submit your portfolio

Task 1.1P Understanding your unit
1.1P - Pass Task

Task 2.1P Basic scripting with P...
2.1P - Pass Task

Task 3.1P Scikit-learn for machin...
3.1P - Pass Task

Task 4.1P Attack Classification u...
4.1P - Pass Task

Task 4.2C Intrusion Detection us...
4.2C - Credit Task

Task 5.1D End-to-end project deli...
5.1D - Distinction Task

Task 7.1P Taxonomy of Attacks, ...
7.1P - Pass Task

Task 8.1P Test your privacy kno...
8.1P - Pass Task

Task 10.1P ACS Report on Priva...
10.1P - Pass Task

The burndown chart shows how much work remains for you to achieve your target grade.

Aim to keep your Complete line close to or ahead of the Target line to keep on track.

There are very few things that I would change in the way I went about this unit. But if I did this unit again, then I would like to compliment my skills by developing a capstone project that summarizes the skills acquired during this unit (I will pursue this as a side project in the future). I would also like to manage my time better.

As an AI professional, this unit has made me aware of the need for conscious efforts to maintain data privacy while developing artificially intelligent algorithms. This will enable me to lead efforts in my workplace into writing data privacy secured solutions and migrating existing solutions towards state-of-the-art solutions.