6COM1031 - Project Planning

Assessment 3 – Project Critique

*“Emotional stylometric analysis using AI, and its applications on cyber-security.”*

Section 1 – Project Description

This project attempts to solve a problem present in many companies, according to the author’s research. He calls the phenomenon “insider threat”, which in short tells us that most of the cyber-attacks to companies are caused by individuals within the company that is attacked rather than external sources. His proposal is to create a system capable of analysing all text communications written by every employee in a company to detect and prevent the possibility of an attack against the company by said individuals. It should be noted however, that despite the title appealing for the cyber-security applications this system would have, it later focuses on using the software on the company’s employees, which could be interpreted a flagrant violation of their privacy.

Said that, the idea sounds reasonable and worthy of studying in more detail, since it could have different applications, rather than just being used to find threats inside a company by analysing the employees’ social media posts and private emails. It could for instance be used to unveil terrorist threats or other sorts of unlawful activities.

As mentioned in the following article[1], "Predicting terrorist events is a dream, but protecting some area by using patterns is a reality. If you know the patterns, you can reduce the risks. It's not about predicting, it's about understanding."

***217 words.***

Section 2 – Evaluation of Project

This report is clear, easy to understand and well structured. It is connected adequately and leads to believe that the Scientific Method was properly executed. A good evidence of this is that the *Research*, *Design*, *Implementation* and *Evaluation* sections are well defined and match what the reader would expect to find inside them, thus making it clear that the author followed a series of steps while working on the project.

The *Research* part of it includes the work done by the author to try to understand the problem and the existent studies on the topic of his project. However, this part of the report feels like it includes sections that don’t belong there, like the *Datasets* or *Testing and Evaluation* sections, where the author doesn’t provide much research or evidence, but instead explains how things are going to be implemented.

The *Design* part follows, and displays the requirements and an outline of the implementation that is going to be done.

The *Implementation* section includes details on the components of the software and how they look like inside.

Finally, the *Evaluation* is quite objective and lets the reader know the parts of the project that could be improved.

Although the solution may at first seem trivial and somewhat simple to understand, it proved to be surprisingly effective for its simplicity, averaging a 28% success rate. The hypotheses were therefore validated. It was also noted that there’s a lot of room for improvement, which by no means signifies that this piece of work was a failed project.

While I don’t agree with the initial interpretation of the problem and purpose of the program, since there is no way that your employees (or ex-employees) would grant their companies permission to access their private data and emails, I do agree with a sentence the author writes by the end of the report, under the *“Future Work”* paragraph. This system has potential to be used with social networks like Twitter for statistical purposes, among others.

***330 words.***

Section 3 – Assessment of Conclusions

As the author states, all of the functional and non-functional requirements were met by the system, which is a good start. The author also seems objective while analysing and giving his opinion on the results after the testing, as it can be seen at the *Table 4* comment, where he explains the reasons for the extraordinary values, which could have led many people on attempting to exaggerate the success of their system.

The *Reflection* section however, is really brief and could be way more extensive given the amount of information present in the whole report. Nevertheless, the *Problems Encountered* section is way too long for the amount of information it provides to the reader, which could be summed up in a few words: *“lack of quality data available”.*

In some cases, though, the excessive objectivity of the author makes the conclusion feel as if it were a statement of the objectives that have been met, rather than the actual author’s impressions and thoughts on the performance and overall opinion on the system created.

As I previously stated, the author never mentions the ethical and privacy issues his idea could face, the most obvious one being to get access to the private data and/or email of employees, ex-employees or even third-parties.

However, as the author does mention under the section “*6.3 Future Work”*, his software could be used with data published on social networks for other purposes, rather than using it on private data for surveillance.

In my opinion, most of the Future Work paragraph could be omitted, since the only relevant part to the question is the last sentence: “This report […] could begin research into creating a large-scale corpus of emotion annotated text.”

The *Conclusions* section was not too extensive nor relevant on the original report and therefore, not much more analysis or criticism can be made.

***308 words.***

Appendix A – Proposed Project Plan

My project is aimed at the development of a Chess game. I wanted to make an outstanding project, including as many of the concepts that I have learnt as possible. Chess is a game I always enjoyed playing, especially when I was a child and always wondered how computers could be better than humans at such complex games where a brute force approach isn’t valid.

The tasks and subsections will be:

* Development of the game. This will include:
  + Java Programming (Java Enterprise Edition): Chosen programming language.
  + User Interfaces: Based on Swing.
  + Design Patterns.
  + Threading.
  + System Tests: Every function and feature must be tested.
* At this point, two players should already be able to play the game in a same computer without being able to perform any illegal moves. The second part will involve:
  + AI Algorithms: Featuring Min-Max, Neg-Max, Alpha-Beta Pruning and MonteCarlo Tree Search.
  + Neural Networks: Optional. Could vastly enhance the AI.
* Although my priority is to finish the points above, these could be optional tasks:
  + Ability to play online.
  + Generation of a database with the history of moves.
  + Distributed Computing: A more advanced version of the AI could be hosted at a server.
  + The game could also have web and/or mobile versions.

***214 words.***

*[1]* [*https://phys.org/news/2017-03-terrorist-behaviors-percent-accuracy.html*](https://phys.org/news/2017-03-terrorist-behaviors-percent-accuracy.html)

## Gantt Diagram:

I currently don’t have the tools for said diagrams installed, so I created a table instead that can be used to generate the actual diagram. It should also be noted that since I’m an Erasmus student, my project is way longer and extensive than most of the ones here. It needs to be like this since I need to present it at my home university at the end of the university year. Also note that in my university, the report doesn’t have as much importance and doesn’t need a whole semester to be written, whereas the actual project does and starts on the first semester, therefore I started with the development about 6 weeks ago.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **#** | **Task Description** | **Duration** | **Start** | **End** | **Predecessors** |
| 1 | First approach to the concept[2] | 2 days | 17/10/17 | 19/10/17 | - |
| 2 | Contact and permissions[3] | 5 days | 20/10/17 | 25/10/17 | 1 |
| 3 | Initial idea design | 3 days | 26/10/17 | 29/10/17 | 2 |
| 4 | Design of the data structures | 6 days | 30/10/17 | 05/11/17 | 3 |
| 5 | Creation of the user interfaces | 15 days | 06/11/17 | 21/11/17 | 4 |
| 6 | Creation of the game algorithms[4] | 35 days | 06/11/17 | 11/12/17 | 4 |
| 7 | AI research phase[5] | 15 days | 21/11/17 | 06/12/17 | 3 |
| 8 | Min-Max implementation | 5 days | 06/12/17 | 11/12/17 | 7 |
| 9 | Neg-Max implementation[6] | 4 days | 12/12/17 | 16/12/17 | 8 |
| 9b | Neg-Max continuation\* | 6 days | 06/01/18 | 12/01/18 | 8 |
| 10 | Alpha-Beta implementation | 10 days | 13/01/18 | 23/01/18 | 9 |
| 11 | MonteCarlo Tree Search[7] | 25 days | 24/01/18 | 18/02/18 | 10 |
| 12 | Online version[8] | 7 days | 19/02/18 | 26/02/18 | 6 |
| 13 | Neural Networks approach[9] | 15 days | 27/02/18 | 14/03/18 | 11 |
| 14 | Distributed Computing[10] | 10 days | 15/03/18 | 25/03/18 | 13 |
| 15 | Finish Report | 10 days | 26/03/18 | 06/04/18 | 14\*\* |
| 16 | Translate and adapt report[11] | - days | 06/04/18 | --/--/-- | 15 |
| 17 | Error correction | 20 days | 26/03/18 | 16/04/18 | 14 |

*\*Christmas break. Some progress and debugging might be done, but I preferred to omit it as it won’t be significant.*

*\*\*It should be noted that not all of the parts of the project MUST be finished. This is just an optimistic approach and represents the maximum number of functionalities that could be done. The report will be written while the development tasks are performed, not left for the last weeks.*

*[2] Choosing the project topic.*

*[3] I had to contact a few people at my home university and here for them to approve my project and give me the green light to start with the development.*

*[4] This is a long and tedious task and involves multitude of tests to ensure every function works as intended and there are no unforeseen edge cases. It also involves a lot of debugging, hence its duration.*

*[5] Choosing and studying the most suitable AI algorithms for the project.*

*[6] Neg-Max is a variation of the Min-Max algorithm, but it might take some time to implement since it’s a new concept for me.*

*[7] This will be the main topic of my research; therefore, it will need quite some time to be finished.*

*[8] It won’t take long presumably, since I’ve done a similar project in the past.*

*[9] This will be an experimental approach and will involve research too. It might not be feasible for the final game, but it’s worth trying.*

*[10] The AI algorithm and the server would be hosted at a server in France.*

*[11] The report will finally be handed in Spanish at my home university. The deadline for the project is the 27th of May, therefore, some features might be incomplete before the hand-in date here in the UK.*