**Group Number: Title of your Projects**

For example: Group 5: Build a recommender systems using collaborative filtering

Your submissions:

* Group number\_Report.pdf
* Group number\_Codes.ipynb (with necessary comments)

Notes

* No extension to the deadline
* Each team can only submit one copy by a single member, just list all of your members in the report
* use RED font for the parts that you revised according to the feedbacks in your presentation

|  |  |  |
| --- | --- | --- |
| First Name | Last Name | Email address |
| Piotr | Rajchel | prajchel@hawk.iit.edu |
|  |  |  |
|  |  |  |

Table of Contents

[**1. Introduction** 2](#_Toc56029461)

[**2. Data** 2](#_Toc56029462)

[**3. Problems and Solutions** 2](#_Toc56029463)

[**4. KDD** 2](#_Toc56029464)

[4.1. Data Processing 2](#_Toc56029465)

[4.2. Data Mining Methods and Processes 2](#_Toc56029466)

[**5. Evaluations and Results** 2](#_Toc56029467)

[5.1. Evaluation Methods 2](#_Toc56029468)

[5.2. Results and Findings 2](#_Toc56029469)

[**6. Conclusions and Future Work** 2](#_Toc56029470)

[6.1. Conclusions 2](#_Toc56029471)

[6.2. Limitations 2](#_Toc56029472)

[6.3. Potential Improvements or Future Work 2](#_Toc56029473)

# **1. Introduction**

Xxxxxxxxxxxx

# **2. Data**

Where did you find the data, how large it is, what are the variables

When it comes to preprocessing the data, my first step before even downloading the data was deciding which columns I needed. The database provides a huge number of rows and columns, but many are for storage purposes, so rather than removing them after downloading the database, I did it ahead of time on the website. I also did not include data below “research grade”, meaning that each observation has been reviewed by peers. These are the columns I chose to leave:

A screenshot of a computer

Description automatically generated with medium confidence

# **3. Problems and Solutions**

Describe the problems to be solved

Note: they are problems, not the solutions. For example, hypothesis testing, ANOVA, classification, clutering, etc, are all solutions

# **4. KDD**

## 4.1. Data Processing

xxxxx

## 4.2. Data Mining Methods and Processes

xxxx

# **5. Evaluations and Results**

## 5.1. Evaluation Methods

xxxx

## 5.2. Results and Findings

xxx

# **6. Conclusions and Future Work**

## 6.1. Conclusions

xxx

## 6.2. Limitations

xxx

## 6.3. Potential Improvements or Future Work

Xxx