Student name: Maxwell Agbana

Minor BDM.

Student number:1660204

Lecturer: Witek Ten Hove

Data mining Portfolio

Contents

[Executive summary 1](#_Toc149421333)

[Introduction 4](#_Toc149421334)

[CRISP model 5](#_Toc149421335)

[Business understanding 5](#_Toc149421336)

[Data understanding 5](#_Toc149421337)

[Data Preparation 5](#_Toc149421338)

[Modelling 6](#_Toc149421339)

[Evaluation 6](#_Toc149421340)

[Naïve Bayes Model 6](#_Toc149421341)

[What I have learnt 7](#_Toc149421342)

[Reflection on Data Mining course 7](#_Toc149421343)

[Appendix 8](#_Toc149421344)

## Executive summary

The project that I am looking to learn and gain a deeper knowledge and understanding of is the use of AI in texts, articles journals. I want to try to make a code that is able to detect if AI came up with passages or if it was self-written by an individual or a group. At the moment, there is no running code that is certified in AI detection however I want to learn from their shortcomings and find out if it is even possible to figure out if AI wrote a passage or not. Furthermore, I also want to see if it is possible for one to camouflage the use of AI within their articles. There are many algorithms that have the ability to detect plagiarism within articles. This is used most notably by schools and universities to control and keep the integrity of students’ assignments. How effective are these algorithms overall though? Over the course of this report, I will put some to the test with articles written by both myself and some retrieved online.

I used python to make a code that code that tries to detect AI usage and plagiarism by identifying if the contents come from the internet or has been written by some sort of AI. This tool can help firms and educational institutions in assessing pieces of texts. Over the course of my research, I used the CRISP model as guidance in not only understanding but learning, testing out and evaluating different models.

1. Linguistic Analysis:

Linguistic analysis remains a fundamental tool for detecting AI-generated content. Common characteristics include unnatural language patterns, a lack of nuance, or inconsistencies in writing. NLP (Natural Language Processing) tools and algorithms, combined with human editorial review, can help identify these anomalies.

2. Data Anomalies:

AI-generated content often relies on large datasets for training. Detecting AI use can involve tracing the origin of the data and looking for inconsistencies or unusual patterns. Anomalies in citations, data sources, or references can be indicative of AI involvement.

3. Algorithmic Footprints:

Various AI models and frameworks leave behind distinctive footprints in generated content. Detection methods include checking for tell-tale signs of specific AI algorithms, like GPT-3, BERT, or Transformer models, which often produce content that is excessively fluent, yet contextually ambiguous.

4. Metadata Analysis:

Metadata, including timestamps and authorship information, can provide clues to the use of AI. Sudden content generation spikes or irregular authorship patterns may indicate AI involvement.

5. Evaluation of References and Citations:

AI-generated content may exhibit unusual citation patterns, such as excessive reliance on obscure sources or a lack of recent references. Verifying the accuracy and relevance of references can help identify AI-generated content.

6. Cognitive Bias and Subjectivity:

AI-generated content may reflect the cognitive biases embedded in the training data. A systematic analysis of content for bias or subjectivity can help pinpoint AI-generated materials, as AI models tend to replicate and amplify the biases present in the data.

Cognitive bias refers to the systematic patterns of deviation from norm or rationality in judgment, often influenced by our individual experiences, beliefs, and heuristics. Cognitive biases can affect how we perceive information, interpret data, and make decisions. Some common cognitive biases include:

Confirmation Bias: The tendency to seek out information that confirms pre-existing beliefs while ignoring or discounting contradictory information.

Anchoring Bias: The inclination to rely too heavily on the first piece of information encountered when making decisions.

Availability Heuristic: Making judgments based on readily available information, which may not accurately represent the overall reality.

Implicit Bias: Subconscious attitudes and stereotypes that affect our judgments and decisions about others.

In the context of AI-generated content, cognitive bias can be manifested in the language used, the examples provided, or the framing of arguments. For instance, an AI model trained on biased data may produce content that inadvertently perpetuates stereotypes, displays political bias, or demonstrates a lack of nuance in its understanding of complex issues.

Subjectivity refers to the quality of being based on or influenced by personal feelings, interpretations, and opinions rather than facts or objective observations. It is a fundamental aspect of human communication and writing, as individuals often express their thoughts, emotions, and perspectives through language. Subjectivity can manifest in various ways, such as:

Personal Pronouns: Frequent use of personal pronouns like "I," "me," and "my" can indicate a subjective perspective, where the author's own experiences and opinions are prominent.

Emotional Language: The use of emotional words and expressions can convey a subjective tone, reflecting the author's emotional state or stance on a particular topic.

Value Judgments: Statements that involve value judgments, such as labelling something as "good" or "bad," are inherently subjective because they are based on individual assessments.

In AI-generated content, subjectivity can be detected by examining the language for signs of personal opinion, emotional tone, or value judgments. AI models that have been trained on data containing subjective content may produce text that reflects these biases and subjectivity, even if it's unintentional.

Subjectivity refers to the quality of being based on or influenced by personal feelings, interpretations, and opinions rather than facts or objective observations. It is a fundamental aspect of human communication and writing, as individuals often express their thoughts, emotions, and perspectives through language. Subjectivity can manifest in various ways, such as:

Personal Pronouns: Frequent use of personal pronouns like "I," "me," and "my" can indicate a subjective perspective, where the author's own experiences and opinions are prominent.

Emotional Language: The use of emotional words and expressions can convey a subjective tone, reflecting the author's emotional state or stance on a particular topic.

Value Judgments: Statements that involve value judgments, such as labelling something as "good" or "bad," are inherently subjective because they are based on individual assessments.

In AI-generated content, subjectivity can be detected by examining the language for signs of personal opinion, emotional tone, or value judgments. AI models that have been trained on data containing subjective content may produce text that reflects these biases and subjectivity, even if it's unintentional.

AI and Cognitive Bias/Subjectivity:

AI models, including natural language processing models, can inadvertently perpetuate cognitive bias and subjectivity if the training data used to build them is biased or subjective. Detecting AI-generated content with cognitive bias and subjectivity may involve assessing whether the content perpetuates stereotypes, lacks nuance, or displays an emotional tone inconsistent with the purported author's style.

Addressing these issues in AI-generated content is essential for ensuring ethical and fair communication. It requires responsible data curation, transparency in AI development, and ongoing efforts to reduce bias and subjectivity in AI models and the content they generate.

7. Style Consistency:

AI models often struggle to maintain a consistent writing style, particularly over lengthy documents. Assessing style inconsistencies, deviations in tone, or shifts in terminology can provide insights into AI involvement.

8. The Human-AI Collaboration Indicator:

While AI tools can assist human authors, certain AI-generated content may still retain subtle human touches. Detecting such cooperation between humans and AI may require deep editorial analysis to unveil AI contributions.

Conclusion:

The detection of AI in reports, articles, and essays is a complex and evolving challenge. Employing a multi-faceted approach that combines linguistic analysis, data examination, algorithmic footprint assessment, metadata scrutiny, and critical evaluation of content is essential. As AI continues to advance, ongoing research and vigilance in this field are crucial to maintaining the authenticity and trustworthiness of written materials. Embracing technology while upholding ethical and editorial standards is the key to navigating the AI-driven information landscape.

## Introduction

The rapid advancement of artificial intelligence (AI) technologies has led to an increasing prevalence of AI-generated content in various written forms, including reports, articles, and essays. Discerning whether AI has been used to create or enhance such content is crucial for transparency, accountability, and maintaining the integrity of information. This report presents a comprehensive analysis of methods and techniques to detect the presence of AI in written materials.

## CRISP model

### Business understanding

The purpose of this research is to try to come up with a model that can detect the use of AI in passages and or pieces of text. Looking at the professional aspect nowadays, AI is at a stage where it has been designed to absorb and utilize human knowledge to its fullest extent. In addition to this, AI is easy to access and is most of the time free for anybody to use. As a result of this, anybody can use articles and writings with slight tweaks that are AI generated and can claim it as their own. This takes away potential opportunities from people that maybe posses the relevant skills that can do the same piece of work without the use or intervention of AI. Using the model that I am going to build, lecturers, schools, companies, and individuals can try to detect if AI has written a piece of text or not. Obviously, this is a very controversial topic it is not highly unlikely or near impossible that one can 100% guarantee that AI has or has not been used in text. But over the course of this journey, I will try to develop something that is as accurate as possible.

### Data understanding

At the moment, a tool called Turnitin is the closest thing to what I want to develop. Turnitin quickly adapted to ChatGPT in order to ensure that students didn’t get a competitive advantage in their pieces of work. It does this by giving a % of how much of the submitted report is from AI or from a source online. (tees, 2013). It does this within 5 minutes of submission and in cases of resubmission, it also tells you how similar it is to the first report. It has a similarity index which is included in the appendix of this report. 0-24% matching text is green/safe, 25-49% matching text is yellow/on the edge, 50-74% is danger zone and 75-100% is a high plagiarism matching text. This is very similar to HandIn at the HAN. TurnitIn has a 98% accuracy rate which is very high in comparison to many other systems and software’s of its kind.

### Data Preparation

Naïve Bayes model

For preparing my data sets, I started with the idea that I would just look online and pick random articles and texts. However, after thinking about it deeper, I figured that it wouldn’t make sense to test my code on those articles because they are already very public and online therefore they can easily detect that the data/ text comes from somewhere. I also had to upload the text on a word document or excel file for python to be able to read it. I tried it on both to see if I would get different results. This also acts as a form of quality assessment. Instead of using texts and articles from online alone, I also attempted it with my own personal assessments. I also decided that I would use my own work to see if it beats the test.

### Modelling

Modelling was the most difficult phase. The aim is to build a model that brings about the best and most accurate output. I also partook a hypothesis testing to look for and test out different kinds of texts from different places. Some online, some my own work. I wanted to see identify how much the results would vary from doing so. I used Chat GPT to start me off with the coding aspect because my coding knowledge was still very minimal. The command I used was “Create a Python code that identifies the use of AI and Plagiarism in a piece of text”. ChatGPT gave me a sample piece of code and I had to do some back and forth trial and errors with the consultation of GPT before the code finally worked.

Naïve Bayes Classifier

The code I ended up writing was a Naïve Bayes code. As stated earlier in this journal, I used my own data and data from external sources as the datasets for my trials. I tried multiple of both variations for control and quality checking reasons. that is used as input, is the original dataset with 150 thousand reviews. For this code we categorized the points into different definitions to make the machine learn what the different points mean.

### Evaluation

### Naïve Bayes Model

When I was done with the code, I tested it out with the various data sets I put forth. I wanted to observe the outcomes and predictions. At the end of the day, even the makers of CHATGPT till this day are not able to 100% detect if it has been used in a piece of text therefore the accuracy of my model cannot be over or understated.

My results/outcomes were quite surprising and interesting to say the least. After the first 3 pieces of texts I used, I had doubts as to whether the code was working/accurate. I used my own reports and pieces of text. I had 3 cases of “This text is Plagiarized” in a row which made me start having my doubts as one of them was inaccurate. I wasn’t sure whether it was because I used sources and references in my piece so I took them out to check if the result of the code would change and it did for 1 out of the 3. Because of this, I did this for all the pieces of text (removing the references) and I got a success rate of more or less 55%. One could argue that for the pieces of text I retrieved from the internet, the accuracy should be certain 100% but this is just a model that doesn’t contain all the data online therefore it cannot guarantee a certain outcome.

## What I have learnt

What I ended up doing is still very basic in comparison to other codes and software’s out there for AI and plagiarism detection. These AI models take time to make and also involve people investing huge sums of money in order to make it work and be available in the market. Over the course of the learning journey, I had to do a lot of research and tests. The research part was quite boring because I’ve been doing a lot of this for the past 3 years at HAN however when I got to the hands-on hard code testing part, It was a little more exciting. I want to work in HR so there isn’t much coding involved in that regard however based on the research that I did, I would not mind pursuing data analytics as a career because I see that in the current labour market, that is one of the best directions to pursue.

## Reflection on Data Mining course

Over the course of this project, I have learnt many things concerning not just data mining but also about making and testing out code. It had been a very long time since I studied anything IT related so it was a good refreshment and was even better to learn new skills that will prove to be of value in the future. Whilst studying and making this Data mining project, I noticed that I didn’t always have to come up with a piece of code from scratch as I had no experience and I wouldn’t learn much by typing up a piece of code but not understanding it. Understanding what I am writing up is crucial to me as that’s how I personally learn. It’s all well and good copying something but if you don’t learn or understand what you are doing then you are wasting your time. Furthermore, it was also fun to use CHATGPT to come up with standard codes and to help find errors in some pieces of code that I had.

Because I worked on my own, I did everything myself with help from outside sources such as Chat GPT and looking at other peoples’ experimental codes on GITHUB. That not only aided but also inspired me as it gave me the confidence to know that a lot of this coding is doable. I needed this confidence because there was a point where I was feeling as though I did not have the ability or capacity to code. I also had to do the business research, data preparation, modelling and deployment on my own which took some time but as stated previously, it was good for my learning.

Furthermore, over the course of Data Mining I also came to the realization that it is a necessity as if you aren’t up to date with the new and emerging technologies, it is your loss. Overall, I am quite proud of what I’ve been able to do because at the start of the course I had doubts as to whether I would be able to do this sort of thing. This project helped me boost my confidence for the other projects that I have to complete.

## Appendix

A yellow bar with black text

Description automatically generated