Logo, company name

Description automatically generated

**George Mason University**

12/9/2022

**Fall**

**2022**

Student Author Name 1

Student Author Name 2

Student Author Name 3

Student Author Name 4

Student Author Name 5

Student Author Name 6

Student Author Name 7

**DAEN 690**

**Project Report**

**Report Title**



Table of Contents

[Abstract 4](#_Toc92633414)

[1 Introduction 5](#_Toc92633415)

[1.1 Background 5](#_Toc92633416)

[1.2 Problem Space 5](#_Toc92633417)

[1.3 Research 5](#_Toc92633418)

[1.4 Solution Space 5](#_Toc92633419)

[1.5 Project Objectives 5](#_Toc92633420)

[1.6 Primary User Stories: 6](#_Toc92633421)

[1.7 Product Vision - Sample scenarios (why would someone want to use this) 6](#_Toc92633422)

[Scenario #1 6](#_Toc92633423)

[Scenario #2 6](#_Toc92633424)

[1.8 Definition of Terms: 6](#_Toc92633425)

[2 Data Acquisition 7](#_Toc92633426)

[2.1 Overview: 7](#_Toc92633427)

[2.2 Field Descriptions: 7](#_Toc92633428)

[2.3 Data Context: 7](#_Toc92633429)

[2.4 Data Conditioning 7](#_Toc92633430)

[2.5 Data Quality Assessment: 7](#_Toc92633431)

[2.6 Other Data Sources 8](#_Toc92633432)

[3 Analytics and Algorithms 9](#_Toc92633433)

[4 Visualization 10](#_Toc92633434)

[5 Findings 11](#_Toc92633435)

[6 Summary 12](#_Toc92633436)

[7 Future Work 13](#_Toc92633437)

[8 Appendix A: Code References 14](#_Toc92633438)

[9 Appendix B: Risk Section 15](#_Toc92633439)

[9.1 Sprint 1 Risks 15](#_Toc92633440)

[9.2 Sprint 2 Risks 15](#_Toc92633441)

[9.3 Sprint 3 Risks 15](#_Toc92633442)

[9.4 Sprint 4 Risks 15](#_Toc92633443)

[9.5 Sprint 5 Risks 15](#_Toc92633444)

[10 Appendix C: Agile Development 16](#_Toc92633445)

[10.1 Scrum Methodology 16](#_Toc92633446)

[10.2 Sprint 1 Analysis 16](#_Toc92633447)

[10.3 Sprint 2 Analysis 16](#_Toc92633448)

[10.4 Sprint 3 Analysis 17](#_Toc92633449)

[10.5 Sprint 4 Analysis 17](#_Toc92633450)

[10.6 Sprint 5 Analysis 17](#_Toc92633451)

[11 References 18](#_Toc92633452)

Table of Figures

**No table of figures entries found.**

Table of Tables

**No table of figures entries found.**

# Abstract

*[Write* ***one*** *paragraph of* ***no more than 300 words*** *that summarizes your project. Here are the typical kinds of information found in most abstracts which you should use as an outline as you develop your abstract.*

1. *The* ***context*** *or* ***background information*** *for your research; the* ***general topic*** *under study; the* ***specific topic*** *of your research.*
2. *The* ***central questions*** *or* ***statement of the problem*** *your research addresses.*
3. ***What’s already known*** *about this question, what* ***previous research*** *was conducted or shown.*
4. *The* ***main reason(s)****, the* ***exigency****, the* ***rationale****, the* ***goals*** *for your research — why is it important to address these questions? Are you, for example, examining a new topic? Why is that topic worth examining? Are you filling a gap in previous research? Applying new methods to take a fresh look at existing ideas or data? Resolving a dispute within the literature in your field?*
5. *Your* ***research*** *and/or* ***analytical methods****.*
6. *Your main* ***findings****,* ***results****, or* ***arguments****.*
7. *The* ***significance*** *or* ***implications*** *of your findings or arguments.*

*Your abstract should be intelligible on its own, without a reader’s having to read your entire paper.*

***NOTE: THE DRAFT ABSTRACT IS WRITTEN IN WEEK 12 OF THE PROJECT AND THE FINAL ABSTRACT COMPLETED IN WEEK 13 OF THE PROJECT****.]*

# Introduction

## Background

*[This section provides 4-5 pages of content for the reader with enough background information about the problem context that allows the paper to be standalone. In other words, assume the reader does not have a background for your project problem and you provide enough content in this section so the reader at a minimum can understand the problem space that you are going to discuss later. Terminology is important and guiding. For example, attempt to minimize the domain vocabulary and when introducing new terms make sure at a minimum, they are defined in section 1.8.*

*Utilize figures, pictures, and tables since visualization are processed faster by the mind.*

*All references throughout the report follow the IEEE Citation Style.]*

## Problem Space

*[This section describes the specific problem that you will attempt to solve completely or part of the problem. N.B., most project scope their solutions to a part of the problem space. Poor project teams usually fail to understand the problem and our eager to start implementing a solution to what they think is the problem. Most problems a very large and understanding the size of the problem mitigates the risk of foolhardy attempts that usually fail. This section should be about 2 pages and should show that the team understands the breadth and dept of the problem space.*

*Even with a well-defined sponsored project that may be scoped it is critical to provide the reader with the whole problem space in a concise an terse description, while providing the reader with a map of what part of the problem you are going to solve. This should allow the team to write the project extensions in the section 7 (Future Work). This is actually a lie since the team will discover many new path and areas of the problem space while working on the project.]*

## Research

*[Document the research the team performed while either developing the solution or attempting to understand a solution path provided by the sponsor. This should be a summary of readings, Internet searches, collaborations, etc. Length of this section can vary but usually around 2-3 pages.]*

## Solution Space

*[Describe your solution approach. High level and provides your reader with an idea of approach.]*

*[Describe the solution space for the reader. For example: “Our system delivers value to its users when it accurately reports veracity scores for submitted articles. Users derive value from these scores when they feel more confident in their chosen news sources or avoid being misled when presented with fake news. We expect our system can help steer users to more authoritative news outlets by altering browsing behaviors.”]*

## Project Objectives

*[What does the team assume it will learn after finishing this project?*

*What does the team assume they will achieve as a solution when they finish this project?*

*What does the team assume it will achieve in terms of understanding about the problem after they finish this project?*

*What does the team assume it will provide in value as a product of this project work to the world, targeted group, etc.?*

*The above questions about the project objectives can be used to develop the primary user stories in section 1.6.]*

## Primary User Stories:

*[This story or stories explicitly state what the project is attempting to address: For example:*

*“Based on the user context and value proposition, we developed the following primary user story to guide our project:*

*“As a User, I want to submit an article to the Veracity System and receive a veracity score to know how much to trust a particular news article.”]*

## Product Vision - Sample scenarios (why would someone want to use this)

* For*:*
* Who*:*
* The*:*
* Is a*:*
* That*:*
* Unlike*:*
* Our product*:*
* Caveats:

### Scenario #1

*[Provide at least two scenarios for this project.]*

### Scenario #2

## Definition of Terms:

*[Glossary can be place here or in an appendix.]*

# Data Acquisition

## Overview:

*[Provide a descriptive overview of your datasets]*

## Field Descriptions:

*[Described your dataset field. Make sure you study the example below and you will more than likely expand these fields:*

* *URL (Type: string) – The web address or Universal Resource Locator for the webpage that contained the news article. This includes the protocol (http or https), host name, and subdomain. Some URLs also include parameters (text following ‘?’) or named anchors (text following a ‘#’). Each URL can only be present once in the database, even if the webpage is not static over time.*
* *Title (Type: string) – The title of the news article as parsed by the Newspaper 3K module. This field may be null (~150 articles in our dataset do not have titles).*
* *Authors (Type: string) –The authors of the news article as parsed by the Newspaper 3K module. This field may be null (~23,000 articles do not have authors) and articles with multiple authors have their names joined with a comma into a single string. This field may also pick up descriptions of the author, including their titles and background.*
* *Publication Date (Type: datetime) – The article publication date and time as parsed by the Newspaper 3K module. The datetime is displayed in ISO 8601 format (YYYY-MM-DD Thh:mm:ss+offset). Publish dates without specified times are assumed to be published at midnight. Publication dates with time information, but without a timezone listing, are assumed to be in Eastern Standard Time. This field is not allowed to be null.*
* *Text (Type: string) – The text of an article as parsed by Newspaper 3K. This field may be null (~8,000 articles do not have text) as some news stories are delivered as only video, audio, or a picture. The mean word count for text is 538.9 across all news sources.*
* *Tags (Type: string) – Article tags as determined by Newspaper 3K. These appear to be important (rare or “topicy”) words taken from the article text, not meta tags contained in the article’s HTML. Multiple tags are concatenated with a comma into a single string.]*

## Data Context:

*[Provide a description of the data context.]*

## Data Conditioning

*[Describe the data conditioning required for each data set.]*

## Data Quality Assessment:

*[At a minimum you must assess your data sets with the following attributes:*

* *Completeness:*
* *Uniqueness:*
* *Accuracy:*
* *Atomicity:*
* *Conformity:*
* *Overall Quality:]*

## Other Data Sources

*[If you are considered other data sources, however, you decided not to use these sources provide some reason why they were not utilized.**]*

# Analytics and Algorithms

*[Provide detailed descriptions of the analytics you used and how this related to your project analyses.]*

# Visualization

*[Describe and show findings and results based on a mix of figures and descriptive text. If you have video, it will be limited to presentation, however, it can also be reference as media file in your Blackboard file exchange.]*

# Findings

*[Summarize your findings and results for the reader.*

# Summary

*[Summarize your overall results for the reader. What did you discover, prove, disprove, etc.]*

# Future Work

*[Critical section! Propose future work or next step(s) for this project.]*

# Appendix A: Code References

*[Provide a GitHub Link and the README.MD content. Do not just provide a link to the GitHub repository but provide a narrative paragraph which introduces the project. This section should mirror the look and feel of a well-documented professional GitHub site.]*

# Appendix B: Risk Section

## Sprint 1 Risks

*[Include the risk table associated with the Sprint. Below the risk table provide a narrative description of how the risks and mitigation plans were identified, what the team got correct, what the team could have done differently, how accurate was the team in identifying the risks, did the team encounter any unanticipated risks, etc. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.]*

## Sprint 2 Risks

*[Include the risk table associated with the Sprint. Below the risk table provide a narrative description of how the risks and mitigation plans were identified, what the team got correct, what the team could have done differently, how accurate was the team in identifying the risks, did the team encounter any unanticipated risks, etc. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.]*

## Sprint 3 Risks

*[Include the risk table associated with the Sprint. Below the risk table provide a narrative description of how the risks and mitigation plans were identified, what the team got correct, what the team could have done differently, how accurate was the team in identifying the risks, did the team encounter any unanticipated risks, etc. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.]*

## Sprint 4 Risks

*[Include the risk table associated with the Sprint. Below the risk table provide a narrative description of how the risks and mitigation plans were identified, what the team got correct, what the team could have done differently, how accurate was the team in identifying the risks, did the team encounter any unanticipated risks, etc. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.]*

## Sprint 5 Risks

*[Include the risk table associated with the Sprint. Below the risk table provide a narrative description of how the risks and mitigation plans were identified, what the team got correct, what the team could have done differently, how accurate was the team in identifying the risks, did the team encounter any unanticipated risks, etc. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.]*

# Appendix C: Agile Development

## Scrum Methodology

*[Provide a narrative of the team efforts in using a scrum methodology for a data analytics engineering project. Describe how easy/difficult was it to adapt to the Scrum methodology. Did the team conduct a daily scrum? If not, how often did the team conduct a scrum. Describe how easy/difficult it was to use the YouTrack tool to manage the project. Don’t be limited to just these questions. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.]*

## Sprint 1 Analysis

*[Provide a narrative of the team’s efforts during this Sprint. Be sure to include – but not be limited to – how the team identified the User Stories, how well the team performed with the various tasks, how easy/difficult it was for the team to manage their activities during the Sprint, what did the team do correct, what could/should the team have done differently, etc. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.]*

## Sprint 2 Analysis

*[Provide a narrative of the team’s efforts during this Sprint. Be sure to include – but not be limited to – how the team identified the User Stories, how well the team performed with the various tasks, how easy/difficult it was for the team to manage their activities during the Sprint, what did the team do correct, what could/should the team have done differently, etc. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.]*

## Sprint 3 Analysis

*[Provide a narrative of the team’s efforts during this Sprint. Be sure to include – but not be limited to – how the team identified the User Stories, how well the team performed with the various tasks, how easy/difficult it was for the team to manage their activities during the Sprint, what did the team do correct, what could/should the team have done differently, etc. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.]*

## Sprint 4 Analysis

*[Provide a narrative of the team’s efforts during this Sprint. Be sure to include – but not be limited to – how the team identified the User Stories, how well the team performed with the various tasks, how easy/difficult it was for the team to manage their activities during the Sprint, what did the team do correct, what could/should the team have done differently, etc. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.]*

## Sprint 5 Analysis

*[Provide a narrative of the team’s efforts during this Sprint. Be sure to include – but not be limited to – how the team identified the User Stories, how well the team performed with the various tasks, how easy/difficult it was for the team to manage their activities during the Sprint, what did the team do correct, what could/should the team have done differently, etc. Think of this writeup as a “lessons learned” that you would like to pass along to any project team thinking of doing a similar project.]*

# References

**There are no sources in the current document.**

*[INSTRUCTIONS: The References section of this document makes use of the Microsoft Word References feature to insert research citations by recording them directly into the document. All citations are to follow the IEEE citation format. Use the Bibliography drop down to have Microsoft Word dynamically create your Works Cited section here in IEEE citation format.*

*To learn more about the IEEE Citation guidelines click on the document links below.*

* [IEEE-Reference-Guide.pdf](https://ieeeauthorcenter.ieee.org/wp-content/uploads/IEEE-Reference-Guide.pdf)
* [IEEE Citation Guidelines2.doc (ieee-dataport.org)](https://ieee-dataport.org/sites/default/files/analysis/27/IEEE%20Citation%20Guidelines.pdf)

*Delete this red text section after you have read and understood the instructions.]*