# George Mason University

# CS 504

# [Fall] [2020]

# **[Project Name]**

**[Draft -> Final] [Report Name]**

**[Team Name]**

**[Team Members]**

# **Table of Contents**

# Abstract

[1-2 paragraph that summarizes your project: a) statement of the problem; b) research methods uses; c) results and findings; and d) conclusions and recommendations.]

# Introduction

## Background and Rationale

## Research

## Project Objectives

## Problem Space

[problem definition]

## Primary User Story (-ies):

[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.””]

## Solution Space

[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.”]

## 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

# Summary

[Summarize your findings and 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 – any code used for the analysis

# Appendix B

Risk Section

# Appendix C

Agile Development

# Appendix D

References