**Project Name:**

Stylometry Analysis of 1800’s Text.

**Group:**

Joseph Allen (solo)

**Abstract/Propose:**

I will perform a stylometry (“the statistical analysis of variations in literary style between one writer or genre and another”) analysis of The Book of Mormon and other writings of Joseph Smith. I will try analyzing how different authors within The Book of Mormon compare to each other and to Joseph Smith.

**Background/Prior knowledge:**

* **What do you already know about the topic, technology, or subject you will be working in. Do you consider your self a Newbie, Beginner, Novice**  
    
  I consider myself in the knowledge of the data I’m going to analyze, in doing data analysis programmatically I feel like a beginner/novice.

**Provide information essential to understanding your project. If applicable, this should include:**

* **Prior work by others - Are you recreating something that has been done before? Are you building on top of others' work?**
  + Others have done similar analysis, but the topic is rather broad in the approach you can take and the different text you can analyze.
* **Prior work by you - Have you already done things in this area, and you are adding to it**?
  + I have not done work in this area.

**Description:**

**Why? – Reason for buy – Why should the who use it? More important than the what? Facts Tell, Stories Sell. Having a good story as to why you are creating the application is always a seller. It also gives you a starts the profile building of your users. You can’t be the sole user, make an aviator of yourself and use it.**

It matters to me personally, It will be interesting to see how Joseph Smith and his different writings compares. I’m approaching it from a faithful angle, so theoretically we should see different writing styles in The Book of Mormon and from Joseph Smith compared to The Book of Mormon. Even if the writing style differences are minimal, there is still faithful perspectives of insights to be gained.

**What?: A Description - In more detail than your abstract, explain what your will project do. What is the solution/features? Does the solution to your problem already exist? If so how is your solution better? Are you re-inventing the wheel? Think of an Info commercial, present your problem, present know solutions and flaws, and convince me why yours is better. What is the real-world impact of your solution? What will prevent them from going back to the old solution or still have the same problems?**

My project will provide more information for The Book of Mormon and also Joseph Smith by analyzing the different writing styles of those authors. My goal isn’t to necessarily prove its truthfulness, but rather to give insight into the text. My goal is to provide insight to those who are approaching it from both a faithful and non-faithful perspective.

**Who?: Target Audience - Describe the intended audience, customer, user of the project. Yes, there is a difference between the customer, who pays for it, and the user, who uses it. What is your primary audience? Who is going to use your solution/program? Are there other markets you could advertise? Will your solution only be in a certain geographic area? community? age range? etc.**

My target audience is people who are interested in The Book of Mormon and Joseph Smith.

**Where? Platform – Where are you going to develop your solution? Where is it going to be used?**

All the information and findings will be posted on Github.

**How? Workflow – How will the who use your solution? How will it work?**

My findings will just be presented in a readable format, anyone who wants to read it should be able to.

**When? SMART Goal - When do you know the project is done? When is good enough, or I have a valid product that I can demo. When do I know I am done with one stage of the project and should move on to the next. This is very similar to Achievable for setting SMART Goals**

There is pretty much an indefinite amount of work I could do on this project. I think a good goal would be to analyze three different “authors” in The Book of Mormon. I want to analyze The Book of Mormon as a whole against other writings of Joseph Smith. I want to do an internal analysis of The Book of Mormon and the authors found (Like Nephi, Mormon, etc.). I lastly want to do an analysis on either the voices (those who speak) within The Book of Mormon, or an analysis of the chapters of The Book of Mormon.

**Significance:**

You can view the significance of my project from two angles. From a faithful perspective it will be another witness to the truthfulness of The Book of Mormon and give more insight to the different authors within it. For a non faithful perspective, a lot of people care about this text and the writer of it. This analysis will provide more context as to why millions of people care about The Book of Mormon.

I think this will be a good project on a resume, I will use the modern tools and techniques for data analysis, which have overlap to the fields I want my career to be in (machine learning and AI development)

**New Computer Science Concepts:**

I’ve had minimal experience with machine learning and stylometry. I want to familiarize myself with it more and also just the tools use to analyze data in programming.

**Interestingness:**

I love data analysis and I also care a lot about The Book of Mormon and Joseph Smith. It will be very fun to work with a text and make it into useable data, while also helping me know more about something I love. Machine learning and other programming data analysis tools allows us to know more about a topic than we ever had before. It can find out important things within data we’d never know without it, and at speeds incomprehensible to most people.

**Tasks and Schedule:**

126 Hours project:

10 Hours: Read about requirements of senior project and write up proposal. Research stylometry techniques and libraries. Research different machine learning models appropriate for stylometry.

10 Hours: Collect all the various text and organize into a format that will be easy for data transformation. If needed house data in a database.

10 Hours: Start experimenting with different data transformation and data structures. Create different sets based off of the different end goals and texts. Begin experimenting with different stylometry libraries and machine learning models.

10 Hours: Do basic stylometry analysis for each text, as in word counts and sentence styles across the different text. Create tables or some other type of data visualization of findings.

10 Hours: Do more advanced stylometry analysis for each text for example, lexical complexity. Compare word usage to other 1800’s popular text. Visualize and explain data.

10 Hours: Develop/train machine learning model and analysis based on authorship of different text within The Book of Mormon. Visualize and explain data.

10 Hours: Develop/train machine learning model and analysis based on The Book of Mormon as a whole and Joseph Smith’s other writings. Visualize and explain data.

10 Hours: Further refine model(s). Research techniques on how to make the models more efficient/accurate. Experiment with different forms of visualization.

10 Hours: Create data/model based off different voices within The Book of Mormon. Fix text data if needed. Visualize and explain data.

10 Hours: Create data/model based off of The Book of Mormon and different text from the 1800’s. Visualize and explain data.

10 Hours: Finalize findings and start creating report for all data and visualization of data.

10 Hours: Finalize report and clean up/fine tune any diagrams and data that need it.

Extra 6 hours: Be added to weeks where they are needed as work is progressing.

**Requirements:**

Shall have a collection and visualization of data for simple stylometry analysis across all text. Success measurement: visualization of data.

Shall have a collection and visualization of data for more complex stylometry analysis across all text. Success measurement: visualization of data.

Shall fit text data to a machine learning model for The Book of Mormon as a whole and Joseph Smith’s other writings, and visualize findings. Success measurement: visualization of data.

Shall fit text data to a machine learning model for the inner voices and authors of The Book of Mormon. Success measurement: visualization of data.

**Stretch Requirements:**

Shall fit text data to a machine learning model for The Book of Mormon and other text that are similar in nature or around same time period (Bible, View of The Hebrews, The Late War, Spauling manuscript)

Shall fit text data of specific voices in The Book of Mormon (Jesus Christ, incarnate and pre-incarnate) to the same voice in the Bible, and other scriptures.