Capstone:

# Letter of Transmittal

Dear Samsung,

As the world’s leading manufacturer in consumer electronics, the importance of understanding the reception of your devices by the customer is tantamount to your continued success. Our firm is proposing a software project to leverage user reviews and sentiment analysis on Reddit forums to understand what customers think about specific products, about Samsung in general, and find receptive subforums for targeted advertising of new products.

# Project Proposal

## Problem Summary

In the recent past, a customer who was interested in buying any form of customer electronics (say a computer mouse) would use Google to search for mouse reviews or about a particular mouse, with the hope of finding a reputable website with reviews or recommendations for the product in question. However, in recent months there has been an emerging trend of dissatisfaction with Google’s search results. Search engine optimization, in tandem with increasingly capable Large Language Models, bolstered by Artificial Intelligence has led to a dramatic loss of trust in Google’s search results. Even once reputable online publications like CNET have recently admitted using AI to generate aggregated “10 Best” lists.

The Washington Post noted in July 2023 that many users were using a so called “Reddit hack” where they would append “reddit” to their Google search (For example instead of “best mouse” they would search “best mice reddit”) in order to get higher quality results. Reddit, which boasts 60 million daily active users, and 100,000 subforums, is a place where people are increasingly flocking to want more human, authentic, and useful results. People go to Reddit to both share and seek out real people and their experiences.

This makes Reddit highly important in two ways. First, Reddit can be a powerful data source for sentiment analysis for various products and services, which can reveal crucial insights about what Samsung is doing right or wrong. This is doubly important, as consumers of our products are themselves doing sentiment analysis in their own intuitive way, looking through user reviews to see which products have the most positive reviews.

Our proposal is an application that will monitor the sentiment of Samsung, and Samsung products across various relevant subreddits, providing real time visualizations in a dashboard.

## Application Benefits

The benefits of this application are numerous. In the most basic sense, this sentiment analysis tool will allow Samsung to monitor their popularity and appeal amongst Redditors, which are a core and highly influential demographic. This can more granularly used to understand what specifically about Samsung or Samsung products creates positive sentiment as well as what shortfalls or lack of features create negative sentiment.

A secondary benefit is identifying advertising opportunities. Reddit has an advertising platform. Understanding which subreddits have a generally high opinion of Samsung products helps target where to advertise certain products as well as what features and capabilities to emphasize.

The same application can also be used to analyze the general sentiment surrounding competitors. For example, in the same subreddits, the same analysis can be run for Dell to see how their products and brand stack up against Samsung’s in the eyes of the public.

## Application Description

There are two main parts to the data product to be developed: one, the machine learning and prediction model (the “back-end”); and two, the Web-based UI to be incorporated into the college’s existing online application process (the “front-end”). The back-end will be developed in Python and will ultimately produce a predictive model that can accurately classify pieces of piano music, matching them to the correct composer. The front-end will be developed using the Charts.js JavaScript framework and will provide a method of interacting with the data.

The posts will be parsed using the Python library TextBlob, which provides an API for natural language processing including part-of-speech tagging, noun phrase extraction, sentiment analysis, classification, translation etc.

MpSentiment Analysis will be implemented using the Natural Language Toolkit, which is based on the VADER sentiment analysis tool. This toolkit is specially designed for social media sentiment analysis.

The application itself will be created using Flask.

## Data Description

Using the Reddit API, the application will be able to access and analyze posts and comments across all of Reddit. However for the purposes of demonstrating this application, we will use a dataset hosted on Kaggle.com that includes posts scraped from various “tech” subreddits (/r/pcgaming, /r/battlestations, /r/mousereview, /r/audiophile, /r/mechanicalkeyboard & /r/buildapcsales)

The dataset includes the following

* Title
* ID
* URL
* Body Text
* Number of Comments
* Total Score (popularity)
* Time Stamp

The data is stored in CSV format. The main limitation of this data is that is a very limited look at previous posts in a small number of subreddits. The value of this application is real time sentiment analysis. The Reddit API accessed via PRAW (Python Reddit API Wrapper) provides post and comment data in much the same format as the .CSV file and can be used to implement the live service application.

## Methodology

## Funding Requirements

## Stakeholder’s Impact

## Developer Expertise

# Technical Proposal

## Problem Statement

## Customer Benefits Projected Timeline Resources and Costs Implementation Plan Evaluation Plan Timeline and Milestones