

Mohsen Jafari – Project Portfolio

Software Engineer, Software Developer, Full-Stack Developer

Tel: (419)378-6913

Email: jafari.mohsen.official@gmail.com

LinkedIn: [linkedin.com/in/mohsen-jafari-mj](https://www.linkedin.com/in/mohsen-jafari-mj)

Introduction

Proficient software engineer with over 15 years expertise in development, specializing in software systems design, optimization, and analysis. Committed to delivering high-quality work efficiently and effectively. Skilled in Software Development Life Cycle (SDLC), Software Engineering (SE) methodologies, and agile practices, with a track record of successful project management. Passionate about advancing skills in machine learning and deep learning to contribute to emerging technologies in computer science. Demonstrated ability to enhance user experience, lead teams, and drive new software product development. Possesses strong analytical thinking skills and a knack for problem-solving, aiming to deliver innovative software solutions. Seeking challenging opportunities to leverage experience and develop high-quality software. Improved comprehensive troubleshooting skills by diagnosing and resolving hardware, software, and network issues at multiple bank branches, ensuring prompt and efficient resolutions.

Projects

1 - Title: Paraillel

Role: Back-End Developer

Technology: React, Python Flask

Description: Paraillel is an AI-driven collaboration platform primarily designed for educators. It helps teachers and school districts enhance collaboration through tools that allow them to share lesson plans, join groups, and coordinate educational strategies. The platform provides analytics to track student performance and helps identify areas where students may need additional support.

Link: [Paraillel.com](https://paraillel.com)

2 - Title: Prioritizing Users' Reviews to Respond in Google Play

Role: Author and programmer


Technology: Machine Learning, Python, Colab, NLP

Description: This project aims to prioritize users reviews on Google Play using Natural Language Processing (NLP) and machine learning techniques to identify which reviews are most important to respond to. Given the large volume of reviews submitted daily, it is a significant challenge for developers to address all of them effectively.

Link: [Prioritizing users review](#)

Classifying Users' Reviews to Respond in Google Play

Mohsen Jafari
MS in Computer Science
Bowling Green State University



SCAN ME

Problem statement

- The number of apps on Google Play has increased.
- Reviews and ratings are important for apps.
- Developers should respond to user reviews.
- Users consider ratings before downloading apps.
- Ratings below 3 stars reduce app downloads.

Research Questions

- What Features of reviews are more important for developer to respond to them?
- What ML models have a better results for classifying reviews?
- How can we improve the classification of reviews?

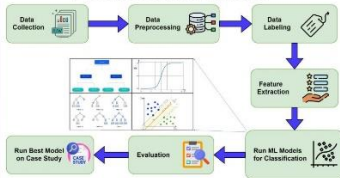
Data Collection

- Dataset has 420,000 reviews and responses
- Dataset has 7 columns
- Dataset is a combination of output of a crawler function and dataset of paper: Automating app review response generation

Data Preprocessing

- Review: remove non-English, break to sentences
- Sentence: break down into individual token
- Word: spell checking, root finding
- Letter: punctuation marks are removed

Architecture of Model



Run Best Models on Case Study

Evaluation

Run ML Models for Classification

Feature Extraction

- Textual Feature: readability, length, complexity, number of nouns, adverbs, verbs
- Semantic Features: sentiment, neutrality, polarity, rating, purpose, title, review inclinations

Classification Results

- Approach #1: Classifying reviews as either requiring a response or not
- Approach #2: Classifying reviews as either high priority or low priority to respond

ML Models

- Logistic regression
- Decision Tree
- Random Forest
- SVM

Evaluation and Results

- Evaluate the models with Accuracy, F1 Score, Recall, Precision
- Approach #1: The best model is SVM with F1 : 0.76
- Approach #2: The best model is logistic regression with F1 : 0.87

Conclusions and Future Work

- Users expect responses from developers, impacting app ratings.
- Developers struggle to address the volume of reviews.
- Use NLP and machine learning to classify reviews.
- SVM has better results for approach #1.
- Logistic regression has better results for approach #2.

Explore automated responses using LLMs.

- Analyze more features.
- Test different ML algorithms.

3 - Title: Predicting Individual Premium Prices: Clustering Subgroups

Role: group member (author and programmer)

Technology: Machine Learning, Python, Colab

Description: Accurate health insurance premium prediction is essential for both insurers and policyholders. This study proposes a two-step approach combining K-means clustering and interpretable regression techniques to improve prediction accuracy and transparency. Results show that this method outperforms traditional models, providing clearer insights into how specific factors influence premium costs.

Link: [Premium Cost](#)

BGSU

Ryan Renken
Ujjwal Kuikel
Mohsen Jafari

QUESTION

To make an accurate prediction on an individual's Premium Price, why train your model with data describing very different people?

METHODOLOGY

- Segment individuals into homogeneous subgroups using K-means clustering algorithm.
- Apply interpretable ML regression models to each cluster separately.
- Evaluate prediction accuracy and determine the best model for each distinct cluster.

Lasso Regression

Feature	Coefficient
Age	18,560.59
Height	0.0
Weight	1,488.41
NumberOfMajorSurgeries	0.0
Diabetes	-584.91
BloodPressureProblems	1,021.18
AnyTransplants	7,043.11
AnyChronicDiseases	3,033.57
KnownAllergies	115.12
HistoryOfCancerInFamily	1,550.49

RF Feature Importance

Feature	Importance Score (approx.)
Weight	0.95
AnyTransplants	0.85
Age	0.75
NumberOfMajorSurgeries	0.65
HistoryOfCancerInFamily	0.55
Height	0.45
AnyChronicDiseases	0.35
KnownAllergies	0.25
Diabetes	0.15
BloodPressureProblems	0.10

TWO-STEP

INTUITION

- By segmenting individuals and treating each subgroup as a separate dataset, models don't need to be as robust in handling large range of complex relationships among variable.
- This approach provides a more interpretable "best" model for a targeted set of individuals

RESULTS

- Lasso Linear Regression was best suited for predicting those with lower costs and less health complications with RMSE 2,952
- Random Forest is best suited for predicting high-cost individuals with more complicated health conditions with RMSE 1,420

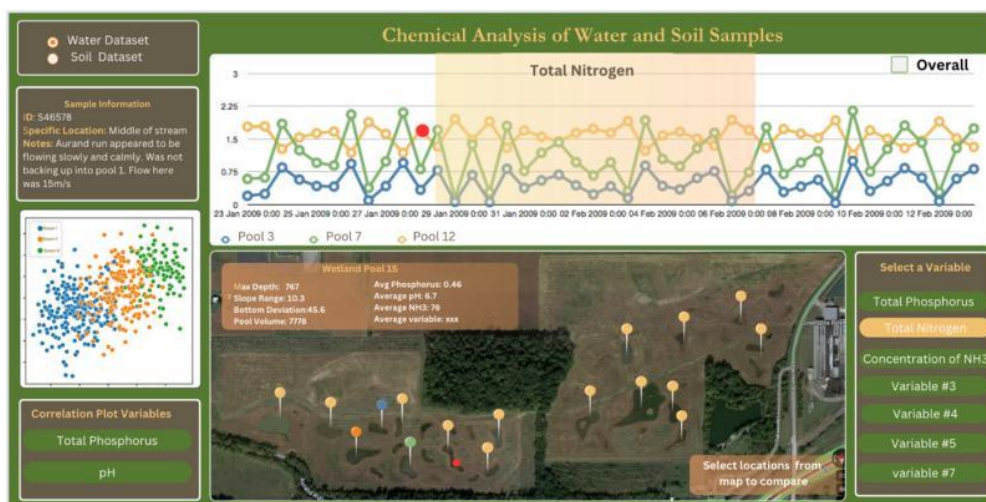
4 - Title: Chemical Analysis of Water and Soil Samples

Role: designer and programmer

Technology: D3.js, html, CSS, GitHub, JavaScript

Description: This is the Term Project for the CS6260 Visualization course. It outlines the implemented design for a visualization tool aimed at facilitating the exploration of scientific data related to soil and water studies. Visualization was implemented as webpage using Html, JavaScript, D3, CSS.

Link: [Vis-project](#)



5 - Title: Latin American History Web Site

Role: designer and programmer

Technology: D3.js, html, CSS, GitHub, JavaScript

Description: I create a website showing the history of some countries in South America. The goal of this project is visualizing the unstructured data in a beautiful way using D3 library and JavaScript.

Link: [Latin American History](#)

6 - Title: Novin Hi-Tech Solution

Role: Team Member (designer and programmer)

Technology: ASP.NET, SQL Server, React, JavaScript, html, CSS

Description: This is the main website of the company I worked for. It features sections for news, products, customers, solutions, and an employee portal.

Link: [Novin Hi-Tech Solutions](#)

7 - Title: Keshavarzi Bank

Role: Team Member (designer and programmer)

Technology: ASP.NET, SQL Server, React, JavaScript, html, CSS

Description: This is the main website of one of our customers. It includes sections for news, products, customers, solutions, and an employee portal.

Link: [Keshavarzi Bank](#)

8 - Title: Preventive Maintenance

Role: Team Leader (designer and programmer)

Technology: C#, .Net, SQL Server

Description: The system was used to schedule the maintenance and repair of the factory machines in such a way that the factories had valuable machines that needed to be repaired and replaced parts at a certain time. It also provided various reports to senior users.

9 - Title: Discrepancy Finder

Role: Team Leader (designer and programmer)

Technology: C#, .Net, SQL Server

Description: This program helps the senior user to check whether all users have withdrawn and deposited the same amount of money in a shared account and if any discrepancy is found, It shows which user made a mistake in his work.

10 - Title: Dynamic Report Maker

Role: Designer and programmer

Technology: C#, .Net, SQL Server

Description: I developed a software called "Dynamic Report Maker" in which End-users could design and make their own reports for total management systems. This software operates like "Crystal Report" in Visual Studio.

11 - Title: Farzan Total System

Role: Designer and programmer

Technology: C#, .Net, SQL Server

Description: I have collaborated in the design and implementation of comprehensive software that included payroll, warehouse, sales, accounting, etc.