Software Requirements Specification

for

Prediction of Adverse Drug Reactions in Pharmacovigilance

Prepared by

19CE018 Vansh Desai and 19CE129 Aman Shah

Charusat University

February 3, 2021

Table of Contents

Revision History

Name	Date	Reason For Changes	Version

1. Introduction

1.1 Purpose

Ultimately, pharmacovigilance is concerned with identifying the hazards associated with pharmaceutical products and with minimizing the risk of any harm that may come to patients. Companies must conduct a comprehensive drug safety and pharmacovigilance audit to assess their compliance with worldwide laws, regulations, and guidance.

Drug safety, is the pharmacological science relating to the collection, detection, assessment, monitoring, and prevention of adverse effects with pharmaceutical products. As such, pharmacovigilance heavily focuses on adverse drug reactions, or ADRs, which are defined as any response to a drug which is noxious and unintended, including lack of efficacy (the condition that this definition only applies with the doses normally used for the prophylaxis, diagnosis or therapy of disease, or for the modification of physiological disorder function was excluded with the latest amendment of the applicable legislation).Medication errors such as overdose, and misuse and abuse of a drug as well as drug exposure during pregnancy and breastfeeding, are also of interest, even without an adverse event, because they may result in an adverse drug reaction.When the pandemic of covid arrived many vaccines were found but each and every vaccines came with some ADR on different people and therefore we need to predict those ADR and for those prediction of ADR we use ML.

1.2 Project Scope

Early innovators in the pharmaceutical industry, alongside health authorities and technology providers, have begun piloting technologies like cognitive computing and ML as an approach to achieve AI to help address data volume challenges. These early experiences can provide us with cautionary insights and important questions that ought to be considered as AI and cognitive computing are entering production systems and are utilized at scale.

When the pandemic of covid arrived many vaccines were found but each and every vaccines came with some ADR on different people and therefore we need to predict those ADR and for those prediction of ADR we use ML

1.3 References

1.Edwards IR, Aronson JK. Adverse drug reactions: definitions, diagnosis, and management. Lancet. 2000;356:1255-9.
2.Lazarou J, Pomeranz BH, Corey PN. Incidence of adverse drug reactions in hospitalized patients: A meta-analysis of prospective studies. Journal of the American Medical Association. 1998;279(15):1200-5.
3.Sultana J, Cutroneo P,Trifiro G. Clinical and economic burden of adverse drug reactions. J Pharmacol Pharmacother. 2013;

2. Overall Description

2.1 Product Perspective

Adverse drug reactions (ADR), defined as 'appreciably harmful or unpleasant reaction[s], resulting from an intervention related to the use of a medicinal product' by Edwards et alaffect millions of people worldwide. ADRs are a huge burden of financial resources and labor. It was estimated by Lazarou et althat more than 100,000 die from ADRs annually.Moreover, Sultana et al found that around $30.1 billion are spent annually on ADRs in the U.S., and nearly half of these costs can be prevented based on a study by Bates et al.The unnecessary high financial costs and labor spent on ADRs provide strong motivation to learn more about ADRs and be able to predict the probability of ADRs accurately.The project's goal is to build predictive models that can predict ADR outcomes given patient demographics and drug prescription information with good accuracy. These models include supervised machine learning models logistic regression, support vector machine, as well as ensemble models random forest and gradient boosted tree. These models can then be used predict outcomes on an individual basis.

2.2 Product Features

- *Accurately measuring the human "gold standard" performance in common PV tasks is critical, allowing us to set benchmarks for evaluating the effectiveness and accuracy of new technologies.*
- *A "human–technology partnership" between human PV experts and cognitive technology may simultaneously offer speed, scale, consistency, and data quality improvements, while maintaining human oversight and control at key decision points.*
- *Thoughtfully training cognitive and AI solutions like ML is critical to their performance. Companies will need to apply their best "business knowledge experts" to reviewing and identifying training data that are representative, diverse, consistent, accurate, and complete.*
- *Both FDA and MHRA expressed interest in exploring use of these technologies in PV, recognizing that they are actively explored by industry.*
- *To demonstrate compliance, new technologies will need to fit within current regulatory and legal frameworks and be equipped with robust quality systems. For audits and inspections, industry will need to be transparent about how such systems are trained and their performance is evaluated.*
- *The panel recognized that the great potential of ML and cognitive computing in PV might best be achieved in partnership with the pharmaceutical industry, IT, and regulatory authorities.*

Read more to review the current state of PV and the major themes discussed during the session

2.3 Operating Environment

Colaboratory, or "Colab" for short, is a product from Google Research. Colab allows anybody to write and execute arbitrary python code through the browser, and is especially well suited to machine learning, data analysis and education. More technically, Colab is a hosted Jupyter notebook service that requires no setup to use, while providing free access to computing resources including GPUs. With Colab you can import an image dataset, train an image classifier on it, and evaluate the model,
all in just a few lines of code. Colab notebooks execute code on Google's cloud servers, meaning you can leverage the power of Google hardware, including GPUs and TPUs, regardless of the power of your machine.
Colab is used extensively in the machine learning community with applications including:

- *Developing and training neural networks*
- *Experimenting with TPUs*
- *Disseminating AI research*

2.4 User Documentation

Through this project we can predict the outcomes of the ADRs. Only we have to do is enter our age, weight, disease , name of medicine, weight and the project will predict the ADR that can be there by usage of that medicine on particular user.

3. System Features

3.1 System Feature 1

3.1.1 Description and Priority

This project presents the results of ML algorithm, a supervised machine learning model for predicting ADR deaths based on patient demographics and drug usage. The model shows good predictive power when trained using balanced samples. Future extensions may include using other models such as ML algorithm to compare the results of the model. Also, when more data become available in the future, it would be interesting to retrain the models and see how the prediction metrics changeS

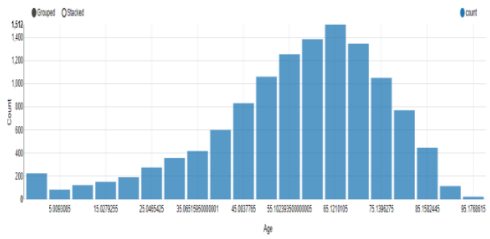

Figure 1: Distribution of patient age.

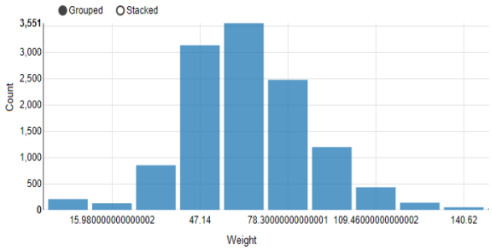

Figure 2: Distribution of patient weight.

Table 1: Metrics of supervised learning models for three most common outcomes.

Outcome	Model	Accuracy	Precision	Recall	F1 Score
Death	Logistic Regression	0.76	0.78	0.71	0.74
Death	SVM	0.69	0.71	0.64	0.67
Death	Random Forest	0.73	0.76	0.68	0.71
Death	Gradient Boosted Tree	0.68	0.75	0.53	0.62
Hospitalization	Logistic Regression	0.75	0.77	0.90	0.83
Hospitalization	SVM	0.73	0.73	0.95	0.83
Hospitalization	Random Forest	0.74	0.73	0.99	0.84
Hospitalization	Gradient Boosted Tree	0.74	0.73	0.97	0.84

3.1.2 Functional Requirements

In this we work with heavy database and in this project user inputs their age, weight, disease, medicine then later on their input is processed in ML and then later on we get output as ADFs of particular data.

4. Other Nonfunctional Requirements

4.1 Software Quality Attributes

- **F1 SCORE**
  Here F1 score can be interpreted as a weighted average of the precision and recall, where an F1 score reaches its best value at 1 and worst score at 0. The relative contribution of precision and recall to the F1 score are equal. The formula for the F1 score is:
  *F1 = 2 * (precision * recall) / (precision + recall)*
- **ACCUARY**
  Accuracy is one metric for evaluating classification models. Informally, accuracy is the fraction of predictions our model got right. Formally, accuracy has the following definition:

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$