CT6039 Dissertation

Student Name: Marcus Abraham

Student Number: s4203822

Course: Computer Science

Submission Date: XX XXX 2024

University of Gloucestershire

**Title goes here**

*Potential title???(Leveraging Ensemble Learning to Enhance Accuracy in Predicting Adverse Reactions to Preclinical Drugs)???*

**Contents**

[**1. Abstract** 2](#_Toc137939224)

[**2. Introduction** 2](#_Toc137939225)

[**3. Literature Review** 2](#_Toc137939226)

[**4. Methodology** 2](#_Toc137939227)

[**4.1 Dataset Collection** 2](#_Toc137939228)

[**4.2 Feature Selection** 3](#_Toc137939229)

[**4.3 Techniques** 3](#_Toc137939230)

[**4.4 Evaluation Metrics** 3](#_Toc137939231)

[**5. Data Pre-processing & Analysis** 3](#_Toc137939232)

[**5.1 Pre-processing** 3](#_Toc137939233)

[**5.2 Analysis** 3](#_Toc137939234)

[**6. Model Development** 3](#_Toc137939235)

[**7. Evaluation** 3](#_Toc137939236)

[**8. Conclusion** 3](#_Toc137939237)

[**9. References** 3](#_Toc137939238)

[**10. Appendix** 3](#_Toc137939239)

# **1. Abstract**

An overview of what the paper contains

# **2. Introduction**

-Start with an overview of the issue presented by ADR’s, and how classical solutions struggle with catching side effects before expensive clinical trials. Highlight both the risks to human life, and costs for companies performing the trials that could have found side effects to developing drugs much sooner.

-State the objectives that this dissertation aims to achieve.

-List the research questions that need addressing.

# **3. Literature Review**

-Gather a mix of around 10 papers/case studies for the review.

-Identify key gaps in the existing research.

# **4. Methodology**

## **4.1 Dataset Collection**

-Describe my datasets and other potential sources of ADR data.

## **4.2 Feature Selection**

-Explain which features I’m selecting from my combined datasets, and why. Cite papers that specifically tie these features to ADR’s.

## **4.3 Techniques**

-Explain each of the techniques and models I’m using in the project.

## **4.4 Evaluation Metrics**

-Explain how I am going to evaluate the success and accuracy of the model.

# **5. Data Pre-processing & Analysis**

## **5.1 Pre-processing**

-Go through how I cleaned, merged and prepared the datasets for analysis. Provide relevant screenshots and link papers which use the same analysis tools I do.

## **5.2 Analysis**

-Give statistics and visual analysis of the data.

-Describe any further feature engineering done.

# **6. Model Development**

-Explain the design and architecture of my model.

-Detail the models training process, hyperparameter tuning and cross validation techniques used.

-Evaluate the models performance via its accuracy, precision, recall, F1 score, and other metrics laid out in the methodology.

# **7. Evaluation**

-Analyse the results I got from my model, and compare them to existing models from the literature review.

-Discuss improvements or limitations that might have held my model back.

-Answer the research questions I wrote in the introduction.

# **8. Conclusion**

-Summarise my findings in the project, and how it has contributed to ADR prediction.

-Identify future improvements to my work, and other potential areas.

# **9. References**

# **10. Appendix**