**High Level Design(HLD)**

Mice Protein

Revision Number :

Last Date of Revision :

|  |  |  |  |
| --- | --- | --- | --- |
| Date Issued | Version | Description | Author |
|  | 1 | Abstract,  Introduction,  General Description | D.Jagannath |
|  | 2 | Design Detail ,KPI,  Deployment | D.Jagannath |
|  | 3 | Final Revision | D.Jagannath |

Contents

Document Version Control…………………………………………………………

Abstract………………………………………………………………………………………………… 4

1.Introduction ……………………………………………………………………….. 5

1.1- Why this High Level Design Document ?……………………….. 5

1.2- Scope………………………………………………………………………………. 5

1.3- Defination……………………………………………………………….. 6

2.General Description……………………………………………………………….. 6

2.1-Product perspective…………………………………………………………………….. 6

2.2-Problem Statement………………………………………………………………………. 6

2.3-Proposed Solution………………………………………………………………………… 6

2.4-Tools Used……………………………………………………………………………………. 7

3.Design Details………………………………………………………………………….. 8

3.1-Process Flow …………………………………………………………………………………… 8

3.1.1-Model Training and Evaluation……………………………………………………….. 9

3.2- Event Log…………………………………………………………………………………….. 10

3.3-Error Handling………………………………………………………………………………. 10

4.Performance…………………………………………………………………………. 10

4.1- Reusabilty………………………………………………………………………………………. 10

4.2- Application Compatibility……………………………………………………………. 10

4.3- Deployment……………………………………………………………………………….. 11

5.KPIs(Key Performance Indicators)………………………………………… 12

6.Conclusion………………………………………………………………… 12

Abstract

The mice protein expression dataset was created to study the effect of learning between normal and trisomic mice or mice with Down Syndrome (DS). The extra copy of a normal chromosome in DS is believed to be the cause that alters the normal pathways and normal responses to stimulation, causing learning and memory deficits. This research attempts to analyse the protein expression dataset on protein influences that could have affected the recovering ability to learn among the trisomic mice. Data mining tasks are employed; Classification analysis via Decision Tree have been proven useful to identify common critical protein responses, which in turn helping in identifying potentially more effective drug targets. Meanwhile, all classification models including the k-Nearest Neighbour, Random Forest, and Naive Bayes have efficiently classifies protein samples into the given eight classes with very high accuracy.

Introduction

* 1. Why this High Level Design Document :

The purpose of this High-Level Design (HLD) Document is to add the necessary detail to the current project description to represent a suitable model for coding. This document is also intended to help detect contradictions prior to coding, and can be used as a reference manual for how the modules interact at a high level.

The HLD will:

* + Present all of the design aspects and define them in detail
  + Describe the user interface being implemented
  + Describe the hardware and software interfaces
  + Describe the performance requirements
  + Include design features and the architecture of the project
  + List and describe the non-functional attributes like:

Security

Reliability o Maintainability

* + - Portability
    - Reusability

Application compatibility o Resource utilisation

Serviceability

1.2. Scope

The HLD documentation presents the structure of the system, such as the database architecture, application architecture (layers), application flow (Navigation), and technology architecture. The HLD uses non-technical to mildly-technical terms which should be understandable to the administrators of the system.

* 1. Definitions :

|  |  |
| --- | --- |
| Database | Collection of all the information monitored by this system |
| IDE | Integrated Development Environment |
| AWS | Amazon Web Services |

2.General Description:

2.1 Product Perspective:

Mice Protein Expression solution is a Machine Learning based model which will help us in prediction of body developments in mice . The data set consists of the expression levels of 77 proteins/protein modifications that produced detectable signals in the nuclear fraction of cortex.There are 38 control mice and 34 trisomic mice (Down syndrome), for a total of 72 mice. In the experiments, 15 measurements were registered of each protein per sample/mouse.The eight classes of mice are described based on features such as genotype, behavior and treatment. According to genotype, mice can be control or trisomic. According to behavior, some mice have been stimulated to learn (context-shock) and others have not (shock-context) and in order to assess the effect of the drug memantine in recovering the ability to learn in trisomic mice, some mice have been injected with the drug and others have not.

2.2 Problem Statement :

Protein Expression classification models are frequently viewed not only as a difficult

task, but also as a classification problem that, in some cases, requires a trade-off

between accuracy and efficiency in analysis validation due to the large amount of data

available.

Expression levels of 77 proteins measured in the cerebral cortex of 8 classes of control

and Down syndrome mice exposed to context fear conditioning, a task used to assess

associative learning.

The aim is to identify subsets of proteins that are discriminant between the classes.

Basically, this is multi-class classification problem

2.3 PROPOSED SOLUTION:

We Build a Machine Learning Model to learn the body developments of mice and down syndrome trsiomic mice which are under development . The model is trained by the data of controlled mice and trisomic mice. Trisomice mice is treated with memantine and saline and control mice also is treated with memantine and saline. Some controlled mice also having some problems in health and some are good and trisomic mice too which is having down syndrome is treated with memantine and saline.

Using trained model we predict the other mice if it is affected with any diseases or having down syndrome according to its health and prediction and the performance of the drug on the mice we used to treat the mice respective treatment.

2.4. Tools Used

    

Python programming language and frameworks such as NumPy, Pandas, Scikit-learn, Matplotlib and seaborn are used to build the whole model.

* + Jupyter Notebook is used as IDE
  + For visualization of the plots, Matplotlib, Seaborn and Plotly are used.
  + AWS is used for deployment of the model.

Mongo DB is used to retrieve, insert, delete, and update the database. GitHub is used as version control system to maintain the code and to store the code.

3.Design Details

3.1-Process Flow : For Finding Health of Control Mice and Down syndrome Mice, we will use a Machine Learning based model to predict the health of mice and respective treatment given to the mice as shown in the below figure.

Proposed Methodology :

Data is fetched from the Mongo DB Database

Data Ingestion

Data Validation

Data Transformation

Model Training

Model Evaluation

Model Pusher

3.1.1- Modal Training and Evaluation :

Dataset is Loaded after data Validation

Data Pre-processing in done in Data Transformation Phase

Model is Build to train and to test the data

Model is Evaluated if the model is accurate or not

Model Pusher to cloud

Prediction

3.2- Event Log :

The system should log every event so that the user will know what process is running internally

Initial Step-By-Step Description:

1. The System identifies at what step logging required
2. The System should be able to each and every system flow
3. Developer can choose logging method. You choose database logging / File logging as well
4. System should not hang even after using so many logging. Logging just because we can easily debug issues so logging is mandatory to do

3.3-Error Handling :

Should errors be encountered, an explanation as to what went wrong . An error as anything that falls outside normal and intended usage.

4.Performance :

The Machine Learning Credit Card Default Prediction solution will be used for to know whether a person able to pay his payment or not by using his past 6 months data . So that Finance company will take necessary action.

4.1-Reusabilty :

The code written and the components used should have the ability to be reused with no problems