Literature Survey

Topic: Visualising and Predicting Heart Diseases with an Interactive Dashboard

Team Members:

1. Pumenitha S T (Team Leader)
2. Geffrey N
3. Shivani T U
4. Karthikeyan P

Table of Contents:

Paper: 1

Title: Intelligent Heart Disease Prediction System Using Data Mining Techniques
Authors: Sellappan Palaniappan, Rafiah Awang
Contribution:

- The prototype for heart disease prediction is developed using three data mining techniques such as Decision Trees, Naïve Bayes and Neural Network.
- Five mining goals are defined based on business intelligence and data exploration.
- The goals are evaluated against the trained models.
- All three models could answer complex queries, each with its own strength with respect to ease of model interpretation, access to detailed information and accuracy.
- Naive Bayes could answer four out of the five goals
- Decision Trees could answer three out of the five goals
- Neural Network could answer two out of the five goals

Methodology:
IHDPS uses the CRISP-DM methodology to build the mining models. It consists of six major phases:business understanding, data understanding, data preparation, modelling, evaluation, and deployment.

1. Business understanding phase focuses on understanding the objectives and requirements from a business perspective, converting this knowledge into a data mining problem definition, and designing a preliminary plan to achieve the objectives.
2. Data understanding phase uses the raw data and proceeds to understand the data, identify its quality, gain preliminary insights, and detect interesting subsets to form hypotheses for hidden information.
3. Data preparation phase constructs the final dataset that will be fed into the modelling tools. This includes table, record, and attribute selection as well as data cleaning and transformation.
4. The modelling phase selects and applies various techniques, and calibrates their parameters to optimal values.
5. The evaluation phase evaluates the model to ensure that it achieves the business objectives.
6. The deployment phase specifies the tasks that are needed to use the models

Paper: 2

Title: Predictive Data Mining for Medical Diagnosis: An Overview of Heart Disease Prediction

Authors: Jyoti Soni,Ujma Ansari,Dipesh Sharma,Sunita Soni

Contribution:

- The different algorithms used in the field of medical predictions are discussed.
- This paper mainly focussed on the prediction of heart disease.
- For heart disease prediction, 15 attributes are discussed using basic data mining algorithms which are incorporated using ANN, Time Series, Clustering and Association Rules, soft computing approaches.
- The accuracy of Decision Tree and Bayesian with genetic algorithm are better in performance than KNN, Neural Networks, etc….

Methodology:

Three different supervised machine learning algorithms i.e. Naive Bayes, K-NN, Decision List algorithms have been used for analysing the dataset.

Tanagra Tool:

- Tanagra tool is used to classify the data and the data is evaluated using 10-fold cross validation and the results are compared.
- Tanagra is a data mining suite build around graphical user interface algorithms.

Decision Tree:

- Decision Tree is a popular classifier which is simple and easy to implement.
- It requires no domain knowledge or parameter setting and can handle high dimensional data.
- The results obtained from Decision Trees are easier to read and interpret.

Naive Bayes:

- Naïve Bayes is a statistical classifier which assumes no dependency between attributes.
- It attempts to maximise the posterior probability in determining the class.
- The advantage of using naive bayes is that one can work with the Naive Bayes model without using any Bayesian methods.

KNN Algorithm:

- The k-nearest neighbou's algorithm (k-NN) is a method for classifying objects based on closest training data in the featurespace and it is a type of instance based learning.

Paper: 3

Title: Effective Heart Disease Prediction using Hybrid Machine Learning Techniques

Authors: Senthilkumar Mohan,Chandrasegar Thirumalai,Gautam Srivastava

Contribution:

- Machine learning techniques were used in this work to process raw data and provide a new and novel discernment towards heart disease.
- The mortality rate can be drastically controlled if the disease is detected at the early stages and preventative measures are adopted as soon as possible.
- The proposed hybrid HRFLM approach is used combining the characteristics of Random Forest (RF) and Linear Method (LM).
- HRFLM proved to be quite accurate in the prediction of heart disease.
- The prediction models are developed using 13 features and the accuracy is calculated for modelling techniques.
- The highest accuracy is achieved by HRFLM classification method in comparison with existing methods.

Methodology:

- The technique of Hybrid Random Forest with Linear Model (HRFLM) is used and it improves the accuracy of heart disease.
- In HRFLM, they used a computational approach with the three association rules of mining namely, Apriori, Predictive and Tertius to find the factors of heart disease on the UCI Cleveland dataset.
- HRFLM makes use of ANN with back propagation along with 13 clinical features as the input.
- The obtained results are comparatively analysed against traditional methods.
- Machine Learning process starts from a pre-processing data phase followed by feature selection based on DT entropy, classification of modelling performance evaluation, and the results with improved accuracy.
- Four types of HRFLM Algorithms are used in to predict heart disease.
    - Algorithm 1: Decision tree-based partition
    - Algorithm 2: Apply ML to find less error rate
    - Algorithm 3: Feature Extraction using less error Classifier
    - Algorithm 4: Apply Classifier on extracted features

Paper: 4

Title: An Analysis of Heart Disease Prediction using Different Data Mining Techniques

Authors: Nidhi Bhatla,Kiran Jyoti

Contribution:

- The objective of this paper is to provide a study of different data mining techniques that can be employed in automated heart disease prediction systems.
- Various techniques and data mining classifiers are defined in this work which has emerged in recent years for efficient and effective heart disease diagnosis.
- The Neural Network with 15 attributes has the highest accuracy rate of 100%
- The Decision Tree with 15 attributes has the accuracy of 99.62%
- The Decision Tree along with genetic algorithm provides 99.2% efficiency
- The Naive Bayes with 15 attributes has the accuracy rate of 90.74%

Methodology:

- The main methodology used is the Data Mining Techniques.
- In addition to Data mining, some other techniques are also used in this method.
- The methodology used are
    - Data Mining and Neural Networks
    - Fuzzy Logic and Genetic Algorithm
    - Data Mining and Supervised Machine Learning Algorithms
    - Data Mining and Genetic Algorithm
    - IHDPS and Data Mining Techniques

Data Mining and Neural Networks

Classification Techniques	Accuracy
Naive Bayes	90.74%
Decision Trees	99.62%
Neural Networks	100%

Data Mining and Supervised Machine Learning Algorithms

Algorithm Used	Accuracy	Time Taken
Naive Bayes	52.33%	609ms
Decision List	52%	719ms
KNN	45.67%	1000ms

Data Mining and Genetic Algorithm

DM Techniques	Accuracy	Model Construction Time	Mean Absolute Error
Naive Bayes	96.5%	0.02s	0.044
Decision Tree	99.2%	0.09s	0.00016
Classification via Clustering	88.3%	0.06s	0.117

IHDPS and Data Mining Technique

DM Techniques	Accuracy
Naive Bayes	86.53%
Decision Trees	89%
ANN	85.53%

Paper: 5

Title: Design And Implementing Heart Disease Prediction Using Naives Bayesian
Authors: Anjan Nikhil Repaka,Ramya G Franklin

Contribution:

- The research focuses on establishing SHDP (Smart Heart Disease Prediction) that takes into consideration the approach of NB (Naive Bayesian) classification and AES (Advanced Encryption Standard) algorithm for resolving the issue of heart disease prediction.
- The Naive Bayes surpasses other techniques in accuracy and it gives an accuracy of 89.77% in spite of reducing the attributes.
- AES yields in high security performance evaluation in comparison to PHEA (Parallel Homomorphic Encryption Algorithm).

Methodology:

- The method used is Naive Bayes Algorithm.
- The steps involved in the process are
    - User Registration and Login
    - Data Collections
    - Classification based on Naive Bayes
    - Prediction
    - Result in PDF format
    - The result is encrypted using AES
    - The encrypted result is given back to database
- The accuracy rate for Naive Bayes is compared with other existing techniques such as SMO (Sequential Minimal Optimization), Bayes Net and MLP(Multi-Layer Perception)

S. No	No. of Techniques	Accuracy (%)	Time(s)
1	Sequential Minimal Optimization (SMO)	84.07	0.02
2	Bayes Net (BN)	81.11	0.02
3	Multi Layer Perception (MLP)	77.4	0.75
4	Navies Bayesian (NB)	89.77	0.01

Accuracy Comparison

S. No	No. of Techniques	Security (%)
1	Advanced Encryption Standard (AES)	98.2
2	PHEA	92.21

Security Comparison