**CHAPTER ONE**

**1.1 INTRODUCTION**

Mental disease diagnosis and particularly Schizophrenia diagnosis is posing a serious task in health industry. The process is one essential and monotonous task for psychiatrists worldwide. Schizophrenia is a serious disorder which affects how a person thinks, feels and acts. Someone with schizophrenia may have difficulty distinguishing between what is real and what is imaginary; may be unresponsive or withdrawn; and may have difficulty expressing normal emotions in social situations (World Health Organization, 2018). And according to Mental Health America (MHA), the vast majority of people with schizophrenia are not violent and do not pose a danger to others. MHA also posited that Schizophrenia is not caused by childhood experiences, poor parenting or lack of willpower, nor are the symptoms identical for each person. Schizophrenia shares symptoms with many other mental illnesses and in some cases does not manifest until several months or even years. Schizophrenia, according to psychiatrists, most commonly strikes between the ages of 16 and 40, and males tend to show symptoms at a slightly younger age than females. In many cases, the disorder develops so slowly that the individual does not know that they have had it for many years. However, in other cases, it can strike suddenly and develop quickly. Experts say Schizophrenia is probably many illnesses masquerading as one.

According to World Health Organization, Schizophrenia affects approximately seventy million of all adults globally. Characterized by delusions, hallucinations, and other cognitive difficulties, Schizophrenia can often be a lifelong struggle (Christian Nordqvist, 2017). A sizable proportion of people with schizophrenia have to rely on others because they are unable to hold a job or care for themselves. Many may also resist treatment, arguing that there is nothing wrong with them. Christian Nordqvist (2017) explained that some patients may present clear symptoms, but on other occasions, they may seem fine until they start explaining what they are truly thinking. He further explained that effects of schizophrenia reach far beyond the patient - families, friends, and society are affected too. Symptoms and signs of schizophrenia will vary, depending on the individual.

Several Schizophrenia Management experts described the disease as characterized by positive and negative symptoms. Positive symptoms include delusion, hallucination, disordered thinking and speech, disordered behaviours. And negative symptoms are social withdrawal, extreme apathy, lack of drive and initiative. There are different subtypes of Schizophrenia namely, paranoid, Hebephrenic, catatonic, childhood schizophrenia, psycho-affective disorder. However the United States’ Diagnostic and Statistical Manual of Mental Disorders 5th edition (DSM-V) in 2013 ruled over all the subtypes and brought them under a single heading “Schizophrenia”. According to America Psychiatric Association (APA), the decision to remove the classification was informed by research experiences that they had “limited diagnostic stability, low reliability, and poor validity”. Some other previously used diagnostic criteria removed from DSM-IV include requirements that a person need to experience bizarre delusions and an auditory hallucination to receive a positive diagnosis. Another important modification was that, to receive a positive diagnosis, a person must have at least one of the following symptoms: hallucinations, delusions, disorganized speech. This decisive policy has made it possible to treat all subtypes of Schizophrenia as one group in this research.

As earlier stated, diagnosis of Schizophrenia is monotonous; confusing because of the winding courses of manifestation the disease is characterized with. Nevertheless recent advances in data management have endowed us with technologies effectively competent at looking into chunks of data, discovering and formulating patterns and relationships capable of giving reasonable predictions. These technologies are known as data mining and machine learning algorithms. Useful knowledge can be exposed from health care system using data mining techniques and can be used to predict the likelihood of patients getting Schizophrenia (Jothikumar, Sivabalan, and Sivarajan, (2015). This kind of systems can serve as a adjunct tool for psychiatrists and medical students to diagnose patients with Schizophrenia. Jothikumar, Sivabalan, and Sivarajan, (2015) further stated in their work published in ARPN Journal of Engineering and Applied Sciences that the knowledge discovery capability of data mining has made the techniques a popular research tool for medical researchers and it is able to predict the outcome of a disease using historical data records of patients .The importance of Schizophrenia prediction system can be viewed from the fact that Schizophrenia is one of the notoriously difficult diseases to diagnose. This is so because its cause and cure are yet known, and courses of manifestation are confusing. The disease shares many symptoms with some other diseases; this makes it a necessity to exclude the possibility of recent episodes being a result of any of the diseases known with similar symptoms. Complete diagnosis process of Schizophrenia involves conducting tests like blood tests to get rid of drug effect possibility; imaging procedures like EEG to exclude brain tumor possibility. These tests cost and need time. Diagnosis process is nested and monotonous that optimizing it through automated processes like data mining processes would serve the health care industry well. The automated predicting system could eliminate conducting any test before diagnosing a patient of Schizophrenia, and could likewise help in early detection of the disease for effective management.

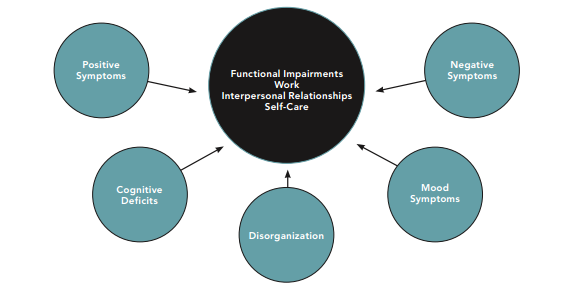
Predicting the outcome of diagnosis is one of the most challenging tasks in data mining. For instance, Schizophrenia patient records contain both redundant and interrelated symptoms, signs, and history data, due to the nature of the disease: “many diseases masquerade as one”. The diagnosis is complex that it takes a lot of processes to arrive at the right decision. Sometimes the decision might even be wrong at the end. This is where data mining intelligent algorithms are useful. They can be used to solve the problems associated with predicting disease outcome using health records dataset involving multiple inputs. Need for fast and accurate diagnosis of Schizophrenia, especially at this time when there is surge in number of reported cases in Nigeria’s psychiatric centers, is a challenge to scientists. Recently, The Punch Newspaper in its October 10th 2018 edition reported the Provost and Medical Director of Neuro-Psychiatric Hospital, Aro, Abeokuta, Ogun State, Nigeria saying that “the medical facility admits between 250 and 300 new patients battling mental illness monthly”. Based on this revelation and coupled with the World Health Organization (WHO) projection of increase in number of persons living with mental illness by 2020, it has become imperative to develop optimized methods of diagnosing and managing Schizophrenia and other mental diseases at reduced cost.

In summary, this study presents essentially a cost-sensitive pruned Decision Tree J48 model that can be used for quick diagnosis of Schizophrenia. The model implements supervised learning procedures with 10-fold Cross-Validation resampling method and utilizes unstructured filter “ReplacingMissingValues” tool in Walkalto Environment for Knowledge Analysis (WEKA) API to replace missing values in the data with the modal values of corresponding features. Feature selection was done using Pearson’s correlation on hot-coded data to detect redundancy in data. Cost Matrix was design to minimize the tendencies of the J48 algorithm to predict False Negative outcomes. This consequently reduces the chance of the model diagnosing a Schizophrenia candidate as free from the disease. The model was found to diagnosis Schizophrenia at 78 per cent accuracy. The model was also repeated on WEKA Explorer, a powerful user-friendly Graphical User Interface of WEKA API.

In addition, the work presents a proposed optimal treatment regime for Schizophrenia management. The treatment regime was carefully designed based on literatures on efficacies and side effects of antipsychotic drugs currently available.

**1.2 SCHIZOPHRENIA DIAGNOSIS PROCESS** (Christian Nordqvist and Timothy Legg 2017).

The figure 1 below shows the features used in diagnosis of Schizophrenia.



**Figure 1: Features of Schizophrenia (Adapted from** [**www.namcp.org/cmeonline.htm**](http://www.namcp.org/cmeonline.htm)**)**

Schizophrenia diagnosis is done by observing the actions of the patient. If the doctor suspects possible schizophrenia, he/she will need to know about the patient's medical and psychiatric histories. Certain tests will be ordered to rule out other illnesses and conditions that may trigger schizophrenia-like symptoms:

1. **Blood tests**: In cases where drug use may be a factor a blood test may be ordered. Blood tests are also done to exclude physical causes of illness.
2. **Imaging studies**: To rule out tumors and problems in the structure of the brain.
3. **Psychological evaluation**: A specialist will assess the patient's mental state by asking about thoughts, moods, hallucinations, suicidal traits, violent tendencies, or potential for violence, as well as observing their demeanor and appearance.

**1.3 SCHIZOPHRENIA DIAGNOSIS CRITERIA**

Patients must meet the criteria outlined in the Diagnostic and Statistical Manual of Mental  
Disorders (DSM). This is an American Psychiatric Association manual used by healthcare professionals to diagnose mental illnesses and conditions. The doctor needs to exclude other possible mental disorders, such as bipolar disorder or schizoaffective disorder. It is also important to establish that the signs and symptoms have not been caused by, for example, a prescribed medication or substance abuse. The patient must have at least two of the following typical symptoms: delusions or disorganized catatonic behavior or disorganized speech or hallucinations negative symptoms that are present for much of the time during the last 4 weeks.  
Patient may experience considerable impairment in the ability to attend school, carry out their work duties, or carry out everyday tasks. Have symptoms that persist for 6 months or more. Table 1 in Chapter three gives summary of the DSM-V criteria for diagnosing Schizophrenia.

**1.4 MANAGEMENT PROCEDURES OF SCHIZOPHRENIA**

International treatment guidelines of organizations, such as World Federation of Societies of Biological Psychiatry (WFSBP), British Association for Psychopharmacology (BAP), National Institute for Health and Care Excellence (NICE) or The Schizophrenia Patient Outcomes Research Team (PORT), stress that pharmacotherapy with antipsychotic medicines, accompanied by adequate psychotherapy is the key to effective schizophrenia treatment. The effectiveness of antipsychotic medications is well established with evidence from numerous clinical trials. Monika Szkultecka-DĊbek et al. (2016) posited that the optimal treatment regime, in terms of antipsychotic selection, dosage, duration, efficacy and tolerability, for each individual patient is less clear, and patient cooperation is a major requirement in treatment.

The management of Schizophrenia should be a comprehensive package that includes the following (Abdul Kadir Abu Bakar, NorainiJali, MohdAminuddinMohdYusof, 2011):

1. Individually-tailored medication
2. Appropriate psychosocial intervention
3. Service level intervention

**1.4.1 TREATMENT BASED ON MEDICATION**

Antipsychotics (APs) are the mainstay of pharmacological treatment in Schizophrenia (Abdul Kadir Abu Bakar, NorainiJali, MohdAminuddinMohdYusof, 2011). Generally APs are divided into Conventional APs (CAPs) and Atypical APs (AAPs). AAPs are used as an option. Examples of CAPs are haloperidol, perphenazine and silpiride. AAPs include amisulpride, olanzapine. Use of AAPs involves close monitoring of the patient. Other commonly used APs are aripiprazole, quetiapine, paliperidone, risperidone, clozapine, Chlorpromazine, fluphenazine, flupenthixol, zuclopenthixol.

**1.4.2 PSYCHOSOCIAL INTERVENTION (**Mayoclinic, 1998-2018**)**

1. Individual Therapy. Psychotherapy may help to normalize thought pattern. Learning to cope with stress and identifying early warning signs of relapse can help people with Schizophrenia manage their illness.
2. Social skill Training. This focuses on improving communication and social interaction and improving the ability to participate in daily activities.
3. Family Therapy. This provides support and education to families dealing with Schizophrenia.
4. Vocational rehabilitation and supported employment. This deals with helping people with Schizophrenia prepare for, find and keep jobs.

**1.4.3 SERVICE LEVEL INTERVENTION (**Mayoclinic, 1998-2018**)**

Service level intervention is currently important in Nigeria in the process of developing more comprehensive hospital-based community psychiatric services and mental health services at the primary care level. It is important the following services be considered for every person with Schizophrenia:

1. Community mental health team to prevent relapse and readmission
2. Assertive community treatment for more difficult cases
3. Supported employment for all who want to work
4. Crisis intervention and home treatment as alternative to acute inpatient care.

**1.5 DATA MINING TECHNIQUES**

Data mining is a promising and relatively new technology. It is a crafty process of exploring, exploiting chunks of datasets, whether homogenous or heterogeneous, with the sole aim of discovering hidden valuable knowledge using various mathematical and statistical computation techniques. According to ZenTut (2018), the needs for standard data mining process increased dramatically. Its applications spread across several disciplines including health care system. Data mining is about how we acquire data, use it to get knowledge, and the use the knowledge to make reasonable decisions, predict the future, understand the past and present, or create new industry/product. It is a process of finding relationship among features defining a dataset. In finding this relationship data mining algorithms are deployed. These algorithms transform information into actionable intelligence. Examples of such algorithms include Decision Trees, Bayesians, Regressions, Neural Networks, Support Vector Machines (SVM), KNN, etc. Many of these algorithms are available in data mining softwares like R, Python, Java, WEKA, etc.

**1.6 STATEMENT OF PROBLEM**

Schizophrenia diagnosis and management involves a nested process of investigations and of course becomes a monotonously rigorous procedure. Psychiatrists have to watch that the observed symptoms are persistent for at least a month as contained in DSM-V classification details and that those symptoms are not as a result of other related diseases nor a behavioral activity(e.g. alcohol drinking) nor a result of a pharmacological side effect. These requirements have made process of diagnosis tortuous and tasking. Unfortunately, the process relies on human judgments which make it prone to errors. Consequent upon this, Psychiatrists need a quicker, effective, and knowledge-based technique to diagnose Schizophrenia for early detection, prompt treatment and effective management of the disease. Similarly, although clinical trials have confirmed efficacy of antipsychotics, selecting optimal treatment regime for individuals is a challenge. It is therefore imperative for scientists to develop quicker ways of diagnosis and to formulate optimal treatment regime that can be adapted for different individuals, especially in the face of World Health Organization (WHO) speculation of increase in mental disorder cases by 2020. This project presents a cost-sensitive J48 model for automated diagnosis and proposes an optimal treatment regime that can be adapted for individual responses.

**1.7 AIMS AND OBJECTIVES**

The aim of this research is to optimize diagnosis and management of Schizophrenia by:

1. Developing a cost-sensitive machine learning model capable of making quick diagnosis for prompt treatment and of minimizing probability of incorrect diagnosis.
2. Presenting an optimal treatment workflow for effective management of Schizophrenia based on results of past clinical trials to establish efficacy of antipsychotics.

The present work is intended to meet the following objectives:

1. Getting Schizophrenia patients’ health records (his and hers).
2. Extracting relevant diagnostic data from the paper files using DSM-V guidance.
3. Converting the data to electronic structured data using Ms. Excel and save as .csv file.
4. Load and select features from dataset on WEKA API
5. Preprocess data on WEKA API software in Python-WEKA wrapper.
6. Building a J48 and a cost-sensitive J48 model that can diagnose schizophrenia.
7. Building predictive models of Bayes Naïve and ZeroR on same datasets.
8. Comparing the performance cost-sensitive J48 to those of other models.
9. Proposing an optimal treatment regime workflow that can be adapted to individual differences.
10. Presenting and analyzing results.
11. Identifying and discussing the research benefits to psychiatrists, patients, and society at large.

**1.8 HYPOTHESIS**

At the completion of the project, the following hypotheses were tested and reported:

1. Cost-sensitive J48 model developed able to diagnose Schizophrenia at reasonable accuracy.
2. The cost-sensitive model able to optimize Schizophrenia diagnosis procedure by saving time and cost associated with diagnosis and management and by minimizing more risky mistakes in diagnosis process.
3. The proposed optimal treatment regime workflow able to optimize management of the disease.

**1.9 SCOPE**

The scope of this project includes:

1. Acquisition of records and psychometric medical data from psychiatric experts of Lagos University Teaching Hospital Nigeria.
2. Preparing obtained medical records data by structuring the unstructured data.
3. Preprocessing of the data: Missing value replacement and feature selection
4. Applying of WEKA Data Mining tools to diagnosis of Schizophrenia.
5. Designing an optimal treatment regime workflow that can be adapted to individual responses.

**1.10 ETHICAL APPROVAL**

This research work was carried out in the Department of Biomedical Engineering, College of Medicine of the University of Lagos. No direct use of human or animal subjects was involved. However, Hospital Diagnosis Records data (151) consisting data of anonymous patients diagnosed of Schizophrenia and those diagnosed of other similar illness by psychiatrists were used to train and test the machine learning models.

**1.11 NOVELTY**

In summary, the project contributes a novel cost-sensitive pruned Decision Tree J48 model that automates diagnosis of Schizophrenia. The model also minimizes, by penalizing through cost matrix design, the chance of making costlier decision errors that can result in diagnosing other diseases for Schizophrenia. This work also contributes an optimal treatment regime workflow that can assist psychiatrists in effective decision making in the management process of Schizophrenia.

**CHAPTER TWO**

**LITERATURE REVIEW**

One of the applications of data mining is medical diagnosis frequently used for research purposes. The research on employing data mining techniques to medical data analysis dated back to nineteen eighties. Data mining diagnosis approach is much more relevant in present times as more diseases emerge and diseases share symptoms, so accurate diagnosis becomes a daunting task. Scientists and medical experts have used the promising tools of machine learning techniques in developing models capable of predicting outcome of disease diagnosis based on features associated with each of those diseases. For instances, DTREE is an expert system that diagnoses DSM-IV Axis I disorders using Decision Tree techniques (Sumathi and Poorna, 2017).Abraham Karplus (2012) compared the performances four different algorithms (Decision Tree, Majority, Nearest Neighbors, and Best Z-Score-a variant of Bayes Naive) on two cancer datasets. The eventual goal of the project in cancer diagnosis is to have a trained machine learning algorithm that, given the gene expression levels or other data from a cancer patient, can accurately predict what type and severity of cancer they have, aiding the doctor in treating it. Masri R.Y. and Jani H.M. (2012) developed a Mental Health Diagnostic Expert system using Rule-Based Reasoning, Fuzzy Logic and Fuzzy-Genetic algorithm. Fuzzy algorithms are complex and difficult to deploy by non-numerate users. Shashikant U. Ghumbre and Ashok A. Ghatol (2012) used India centric dataset for Heart disease diagnosis. The correct diagnosis performance of the automatic diagnosis system was estimated by using classification accuracy, sensitivity and specificity analysis. The study showed that, the SVM with Sequential Minimization Optimization learning algorithm with accuracy of 92 per cent had better choice for the medical disease diagnosis application. TahaSamadSoltaniHeris, MostafaLangarizadeh, ZahraMahmoodvand, Maryam Zolnoori (2013) also developed models for skillful diagnosis of Asthma using machine learning algorithms: Support Vector Machines, Forest Tree and K-Nearest neighbours. The study was conducted on a dataset consisting of 169 asthmatics and 85 non-asthmatics visiting the Imam Khomeini and MasseehDaneshvari Hospitals of Tehran. The algorithms of k – nearest neighbours , random forest , and support vector machine, together with pre – processing and efficient cross folds training were implemented on this dataset ,and the degrees of accuracy and specificity of the system used in the study were calculated compared with each other and with those of previous research. However, although the models achieved high accuracies in the range 96-100 percent, the models are difficult to use as they were implemented in Python language. A psychiatrist not proficient in coding cannot utilize the models. The decision procedures of algorithms are difficult to explain to somebody not sound in Mathematics. In the same vein, Kipli et al (2013) applied Data mining algorithms on Brain Imaging data to classify mental disorders. Imaging data are not easy to acquire and costly. In similar manner, Sarina J. et at (2013) classified Schizophrenia using MR imaging data; Stefan P. Koch et al (2015) proposed a model for diagnostic classification of schizophrenia patients on the basis of Regional Reward-related fMRI signal patterns. DabekFilip et al (2015) employed Neural Network based model for psychological conditions. Swapna G., Vinayakumar R., Soman K.P (2018) employed long short-term memory (LSTM), convolutional neural network (CNN) and its combinations for extracting complex temporal dynamic features of the input HRV data. These features are passed into support vector machine (SVM) for classifiation. They obtained the performance improvement of 0.03% and 0.06% in CNN and CNN-LSTM architecture respectively compared to our earlier work without using SVM. The classifiation system proposed can help the clinicians to diagnose diabetes using ECG signals with a very high accuracy of 95.7%. Krishnaveni et al. (2016) employed J48 and Naïve Bayes algorithm to diagnose and evaluate the Attention Deficit Hyperactivity Disorder. Miseon S. et al. (2016) developed a machine-learning-based diagnosis of Schizophrenia using combined sensor-level and source-level EEG levels. Dinu A.J., Ganesan R1, Felix Joseph and Balaji V (2017) developed an automatic system for measurement of Joint Space Width (JSW) in hand x-ray images of patients suffering from Rheumatoid Arthritis (RA). In the work an approach was proposed for the automatic quantification of radiographic changes in rheumatoid arthritis  
by measuring two indicators for disease progression. Based on a hand radiograph, bone positions and contour delineations were determined by the algorithm. Subsequently, joint space widths  
were measured and erosions on the contours of bones were detected. The work aimed at a quantitative assessment of RA progression that is more sensitive and reproducible than manual  
approaches currently in use. Babita and Depika (2017) used J48 and Case-Based Reasoning for the diagnosis of EEG-based diseases. Shaoqiang et al (2017) used deep-learning method for recognition of early-onset schizophrenia. Similarly, Fahad Saeed et. al. (2018) developed a machine learning algorithm that allows classify fMRI ADHD scans from normal healthy brain scans withoutusing any demographic information. The technique was based on computing similarity between two multivariate time series along with k-Nearest-Neighbor classifier. They designed a model selection scheme called J-Eroswhich is able to pick the optimum value of k for k-Nearest-Neighbor from the training data. The results show a 20% increase in accuracy, with better sensitivity and specificity, as compared to the state of the art algorithms in classifying ADHD using open-data ADHD-200. Sumathi and Poorna (2017) designed an ensemble of Naïve Bayes classifier to predict social and communication deficiency among children. In the work, a clusterer was used along with Bayes classifier for better accuracy. Lebedev, Westman et al (2014) used Random Forest Ensembles for detection and prevention of Alzeimer’s disease. Lee (2017) exploring the performance of Stacking Classifier to predict Depression among the Elderly. Husain, Xin, and Jothi (2016) predicted anxiety among women using Random Forest approach. Abou-Warda, Belal, El-Sonbaty,Darwish (2017) developed a Random Forest Model for diagnosing mental disorders. Thongkam, Sukmak, Mayusiri (2016) developed a Decision Tree Model that could predict Schizophrenia patients’ risk of Readmission in the Long and Short terms using social-demographic and clinical characteristics data. Monika et al (2016) researched on Treatment Patterns of Schizophrenia based on data from seven central and Eastern European Countries. Also, Reetu and Narender (2015) developed a J48- based model for Liver Cancer diagnosis. Vijayarami, Dhayanand applied SVM and Naïve Bayes algorithms for Liver Diseases prediction. Neelamegam and Ramaraj (2013) researched into various types of Classification algorithms in use in Data Mining. Recently, Bo Cao at the University of Alberta’s Department of Psychiatry, with the collaboration of Xiang Yang Zhang at the University of Texas Health Science Center at Houston (2018) used a machine-learning algorithm to examine functional magnetic resonance imaging (MRI) images of both newly diagnosed, previously untreated schizophrenia patients and healthy subjects. By measuring the connections of a brain region called the superior temporal cortex to other regions of the brain, the algorithm successfully identified patients with schizophrenia at 78 per cent accuracy. It also predicted with 82 per cent accuracy whether or not a patient would respond positively to a specific antipsychotic treatment named risperidone. In the same vein, Lei Chu, Robert Qiu, Haichun Liu, Zenan Ling, Tianhong Zhang and Jijun Wang (2018) proposed a new individual recognition schemes based on spatio-temporal resting state Electroencephalography (EEG) data. Instead of using features  
derived from artificially-designed procedures, modified deep learning architectures which aim to automatically extract an individual’s unique features are developed to conduct classification. Our designed deep learning frameworks are proved of a small but consistent advantage of replacing the softmax layer with Random Forest. Additionally, a voting layer was added at the top of designed neural networks in order to tackle the classification problem arisen from EEG streams. Lastly, various experiments were implemented to evaluate the performance of the designed deep learning architectures; Results indicated that the proposed EEG-based individual recognition scheme yields a high degree of classification accuracy: 81*.* 6% for characteristics in high risk (CHR) individuals, 96*.*7% for clinically stable first episode patients with schizophrenia (FES) and 99*.*2% for healthy controls (HC).

While all the above models have been efforts towards solving problems in various capacities and approaches, the sources of data for some of the models are a concern, especially in resource-challenged societies. For instance, fMRI, EEG data are undoubtedly effective representations of patho-physiological conditions of mental disorders but they are costly and require time for acquisition and processing. By and large equally effective knowledge-based models for Schizophrenia diagnosis could be built using more available ICD-10 or SDM-5 based retrogressive diagnostic data obtained by local psychiatrists rather than costly and time-consuming imaging data. Furthermore, many of past studies are not focused on specifics rather spectrum of mental disorders. Fortunately, J48 (ranked no 1 in the top 10 Algorithms in Data Mining paper published by Springer LNCS, 2008) and Naïve Bayes data-mining algorithms are popular effective predictors that can also be extended to Schizophrenia diagnosis as used by Krishnaveni (2016) in his work on Attention Deficit Hyperactivity Disorder. WEKA offers development environment and robust Graphical User Interface as an added advantage.

**CHAPTER THREE**

**3.0 MATERIALS AND METHODOLOGY**

**3.1 DATA ACQUISITION**

Data used for the project were collected from psychiatrists of the Lagos University Teaching Hospital Lagos Nigeria. Dataset consists of 151 health records (his & hers) of patients reported between years 2013 and 2018 inclusive. 105 records are of positively diagnosed Schizophrenia patients, and 46 are of those diagnosed otherwise to serve as controls in the project. Unavailability of electronic health records was a challenge. Availability of data and time constraints are factors considered in deciding number of records reviewed. The dataset has 38 raw attributes including CLASS column. However, only 33 attributes based on DSM-V specifications for definition of Schizophrenia were used in model building with CLASS as labels excluding Year of Patient Report, Age, Sex, and Diagnosis. Table 2 in section 3.2 gives details of DSM-V criteria for Schizophrenia diagnosis. The attributes used were selected to reflect the criteria of Schizophrenia diagnosis as stated in DSM-V documentations. The attributes are listed in Table 1 below. Term represent each feature in WEKA API.

**Table 1 shows Representation of features in WEKA API, Description and Values**

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **FEATURE** | **DESCRIPTION** | **VALUES** |
| **1** | **Y\_O\_REP** | **Year patient reported in Hospital** | **Year e.g. 2017** |
| **2** | **AGE** | **Age of patient** | **Age e.g. 32, 23** |
| **3** | **SEX** | **Sex of patient** | **MALE, FEMALE** |
| **4** | **OCCUP\_HX** | **Occupation History** | **unemployed, occupation** |
| **5** | **MAR\_STA** | **Marital status** | **married,single, divorced, widow** |
|  | **FEATURE** | **DISCRIPTION** | **VALUES** |
| **6** | **DUR\_EPIS** | **Episode Duration(length of time the patient has suffered the symptoms** | **time in months** |
| **7** | **P\_PXY\_HX** | **Past Psychiatric History** | **e.g rape, mental illness, etc** |
| **8** | **P\_MED\_HX** | **Past Medical History** | **No, disease suffered in past(eg diabetes)** |
| **9** | **FAM\_P\_HX** | **Family Psychiatric History** | **Yes, No** |
| **10** | **P\_SOC\_HX** | **Past Social History** | **Yes, No** |
| **11** | **P\_SEX\_HX** | **Past Sexual History** | **normal, experience(e.g masturbate, gonorrhea,etc)** |
| **12** | **FOR\_HX** | **Forensic History** | **No, Yes** |
| **13** | **PREMOB\_HX** | **Pre-morbid History** | **Normal, introvert, extrovert, melancholic** |
| **14** | **MSE** | **Mental State Examination** | **Kempt, unkempt, poor eye contact, restless** |
| **15** | **SPEECH** | **Speech Status** | **Normal, reduced volume, mute, slurred, decreased tone, irrelevant, incoherent** |
| **16** | **MOOD** | **Mood of the patient at the time of report** | **Euthemic, neutral, happy, relaxed, fine/ok, worried, sad, irritable** |
| **17** | **AFFECT** | **Affect of the patient at time of report** | **Depressed, reactive, blunt, restricted, congruent, abnormal** |
| **18** | **TH\_FORM** | **Thought Form at time of report** | **Logical, abnormal** |
| **19** | **TH\_STRM** | **Thought Stream at time of report** | **Reduced, normal, increased** |
| **20** | **TH\_CONTENT** | **Thought content at time of report** | **Persecutory delusion, auditory hallucination, normal, obsession, grandiose delusion, disorder** |
|  | **FEATURE** | **DESCRIPTION** | **VALUES** |
| **21** | **TH\_POSSESSION** | **Thought Possession at**  **time of report** | **Impaired,**  **Normal** |
| **22** | **PERCEP** | **Perception at time of report** | **No, Auditory Hallucination, visual hallucination, tactile hallucination, olfactory hallucination, preoccupation** |
| **23** | **ORIENT** | **Time, Place and Position Orientation at time of report** | **Oriented in TPP, no** |
| **24** | **ATTEN** | **Attention status at time of report** | **Rousable, poor** |
| **25** | **CONC** | **Concentration status at time of report** | **Good, reduced, poor** |
| **26** | **MEM\_IR** | **Immediate Recall status at time report** | **Good, fair, poor** |
| **27** | **MEM\_ST** | **Short-term Memory status at time report** | **Good, fair, poor** |
| **28** | **MEM\_LT** | **Long-Term Memory status at time report** | **Good, fair, poor** |
| **29** | **INT\_GFK** | **Intelligence Test of General Fund of Knowledge** | **Good, fair, poor** |
| **30** | **INT\_S\_A\_D** | **Intelligence Test of Similarity and Difference** | **Good, fair, poor** |
| **31** | **INT\_CAL** | **Intelligence Test of Arithmetic** | **Good, fair, poor** |
| **32** | **INT\_PROV** | **Intelligence Test of Proverbs** | **Good, fair, poor** |
| **33** | **JUDGMT** | **Intellectual Judgment status at time of report** | **Good, poor** |
| **34** | **INSIGHT** | **Insight status at time of report** | **Good, partial ,poor** |
| **35** | **PSE** | **Physical State Examination status** | **Good, normal, pale** |
| **36** | **EEG** | **Electroencephalogram (to exclude brain tumor possibility status).** | **Normal, altered** |
| **37** | **DIAGN** | **Result of diagnosis** | **Diagnosis Result e.g. Bipolar disorder** |
| **38** | **CLASS** | **Classes of instances** | **SCHIZ, OTHERS** |

**3.2 DSM-5 SCHIZOPHRENIA DIAGNOSIS CITERIA/ATTRIBUTES**

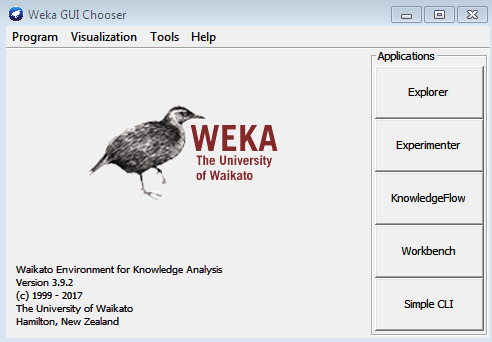
DSM means Diagnostic and Statistical Manual of Mental Disorders. It is documentation prepared and published by the American Psychiatric Association and covers all categories of mental health disorders for both adults and children. It is used by clinicians and psychiatrists to diagnose psychiatric illness. DSM is widely used in the United States for psychiatric diagnosis, treatment recommendations, and insurance coverage purposes (Steven Gans, 2018). The current version of DSM documentation is DSM-V which was released in 2013. Table 2 below shows the criteria for diagnosis of Schizophrenia as defined in DSM-V.

**Table 2 shows DSM-5 Criteria for Schizophrenia Diagnosis**

|  |  |
| --- | --- |
| **Delusion** |  |
| **Hallucination** |  |
| **Disorganized speech** |  |
| **Grossly disorganized or catatonic behavior** |  |
| **Diminished emotional expression or avolition (Negative symptoms)** |  |
| **Social/occupational dysfunction** |  |
| **Disturbance Duration** | **At least six months** |
| **Schizoaffective and mood disorder exclusion** |  |
| **Substance/general mood condition exclusion** |  |
| **Relationship to Global Developmental Delay or Autism Spectrum Disorder** | **Schizophrenia is diagnosed if delusion or hallucination is observed.** |

**3.3 WEKA TOOLS**

WEKA means Waikato Environment for Knowledge Analysis. WEKA is a data mining environment developed by the University of Waikato in New Zealand. WEKA is open source software available under the GNU (General Public Licenses). It provides the user graphical interface for easy access to inbuilt functionalities; also it is a collection of algorithms for data analysis and visualization tools. Using WEKA the algorithms are applied directly to a dataset or called from own Python code using WEKA API in Python-WEKA wrapper. Also, the new machine learning schemes can be developed with this package or software (MohdFauzi bin Othman et al., 2007). Applications written using the WEKA tool can be run on any computers with a Web browsing capability. For the processing of data in WEKA, data is converted into the ARFF (Attribute Relation File Format) of CSV format. It describes lists of instances sharing sets of attributes. WEKA has several graphical user interfaces: Explorer, Experimenter, KnowledgeFlow, WorkBench, Simple CLI but the main graphical user interfaces is Explorer that enables easy access to essential functionality. Figure 2 below shows WEKA GUI Chooser.



**Figure 2 Shows WEKA GUI Chooser**

**3.4 J48 DECISION TREE CLASSIFIER**

J48 classifier is a simple C4.5decision tree induction algorithm for classification. For generating decision tree, a well known tree-growing algorithm that is based on univariate splits is ID3 (Quinlan, 1986), and its extended version is called C4.5 (Quinlan, 1993). The feature that are added into J48 are decision tree pruning, accounting for missing values, continuous attribute value ranges, derivation of rules, etc. J48 is an open source java implementation of the C4.5 which is used in WEKA (International Journal of Recent Scientific Research Vol. 6, Issue, 6, pp.4809-4813, June, 2015). J48 was ranked No 1 in the top 10 Algorithms in Data Mining in a paper published by Springer LNCS in 2008.

**3.4.1 THEORY BEHIND DECISION MAKING OF J48 CLASSIFIER**

Decision trees are learned from training data. Each data item consists of a set of features describing an object and the class of the object. Decision trees are recursively built beginning with the topmost node by:

1. Computing the best test for the current node according to some splitting criterion.
2. Creating a subnode for each possible outcome of the test, and
3. Recursively expanding each subnode in the same way until a given stopping criterionis satisfied.

Usually, the decision tree is afterwards simplified (pruned) in order to avoid overfitting of the training data. The test of the topmost node divides the training data into two subsets. One subset contains the elements which pass the test, the other contains the elements which fail the test. During the induction of the tree, the two subsets are passed on to the *‘*yes’and *‘*no’subnodes respectively. The feature tests of the subnodes further subdivide the data subsets, and so on. The majority class of the data which reaches a terminal node becomes the result class of that node. **3.4.1 .1 The Splitting Criterion**

The best test for a node is selected according to the splitting criterion. A frequently used splitting criterion is the information gain. It is the difference between the entropyof the data set at the current node and the entropy in the two subsets induced by the test. The entropy of a data set measures to which degree the data is scattered over several classes. If a data set is pure, i.e., if all elements belong to the same class, the entropy is 0. If half of the data belongs to class A and half of the data to class B, the entropy is 1. The entropy is defined by the formula:

**H(p) = − p(c) log2p(c)**

where p(c) is the relative frequency (empirical probability) of class c in the data set, i.e., the frequency of class c divided by the size of the data set.

The information gain is defined as follows:

**G = H(p) – w1H(p1) – w2H(p2 )**

where p(c) is the relative frequency of class c in the current data set, p1(c) and p2(c) are the relative frequencies of class c in the two subsets, and **w1** and **w2** are the proportions of data in the first and second subset.

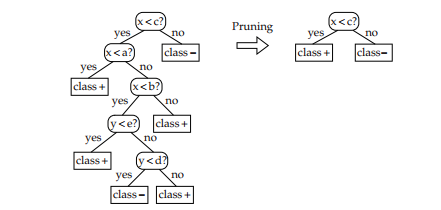
**3.4.1.2 Stopping Criterion**

The recursive expansion of the decision tree is stopped if either the data subset is pure or if all data items have the same feature representation. The latter case occurs when the data contains objects with identical feature values, but different classes. When such contradictory class assignments exist, there is no decision tree which correctly classifies all the training data. Sometimes these stopping criteria are augmented by other criteria which may terminate the induction process earlier, such as:

* 1. The size of the data set being below a certain threshold
  2. The value of the splitting criterion for the best test being below a threshold.

**3.4.1.3 Decision Tree Pruning**

Decision trees which are grown to their maximal size as described above tend to overfit the training data. Overfitting occurs when the classification of the decision tree depends on accidental properties of the training data. Overfitting is a problem because it leads to errors on new data. In order to avoid overfitting, most decision tree learning algorithms add another step which simplifies the decision tree with pruning. Pruning identifies irrelevant feature tests and replaces the corresponding non-terminal nodes with terminal nodes. Critical-value pruning  
(CVP) (Mingers 1987) prunes at a node if the score of the splitting criterion (information gain, gain ratio, or other) is below a given threshold. The pruning proceeds bottom-up, and it only considers nodes whose subnodes are terminal nodes or pruned nodes. The size of the pruned tree depends on the threshold. Higher thresholds lead to smaller trees. In order to determine the optimal threshold, the decision tree is pruned with different thresholds. The pruned trees are evaluated on test data by computing the classification accuracy, which is the proportion of correctly classified test items. The tree with the highest accuracy is selected. It is important that the data used for this evaluation is fresh data which was not used to induce the tree. Otherwise, the tree will not be pruned because the full tree achieves the highest accuracy on the training data. If only a fixed amount of training data is available for tree induction and pruning, part of the data has to be set aside for pruning before the tree is induced. The reduced error pruning (REP) method invented by Quinlan (1987) also requires separate data for pruning, which is here directly used to decide which nodes to prune. A node is pruned if the number of errors on the pruning data is not increased by the pruning, i.e., if the total classificatioln error of the subnodes is at least as high as the classification error of the node after pruning. Again, nodes are only pruned if all non-terminal subnodes have been pruned before. Figure 3 below shows a Decision Tree being pruned.



**Figure 3 shows a typical Decision Tree Being Pruned**

**3.5 ZeroR CLASSIFIER**

A baseline classification uses a naïve classification rule such as:

1. Base Rate: Accuracy of trivially predicting the most frequent class. ZeroR classifier in WEKA always classify to the largest class i.e. according to the prior.
2. Random Rate: Accuracy of making a random class assignment. It might apply prior knowledge to assign random distribution.
3. Naïve Rate: Accuracy of some simple default or pre-existing model.

ZeroR gives a baseline accuracy that must be always checked before choosing a sophisticated classifier. Its Accuracy is also known as null rate.

**3.6 NAÏVE BAYES CLASSIFIER**

It is a classifier that works based on Bayes Theorem

P(A|B) =

Bayes Theorem tells how often A happens given that B happens, written as P(A|B), when we know how often B happens given that A happens, written as , and how likely A and B are on their own written as and respectively.

Naïve Bayes classifier calculates the probabilities for every factor, then it selects the outcome with highest probability. The classifier assumes the features are independent and hence the word naïve. And even with this, it is a powerful algorithm used for many classification problems including medical disease diagnosis.

**3.7 PEARSON’S CORRELATION**

Pearson's correlation coefficient is the [covariance](https://en.wikipedia.org/wiki/Covariance) of the two variables divided by the product of their [standard deviations](https://en.wikipedia.org/wiki/Standard_deviations). The form of the definition involves a "product moment", that is, the mean (the first [moment](https://en.wikipedia.org/wiki/Moment_(mathematics)) about the origin) of the product of the mean-adjusted random variables; hence the modifier product-moment in the name. The coefficient of correlation when applied to two variables X and Y is given by:

rX,Y =

where xi is instant value of X; x is mean value of X; yi is instant value of Y; and y is mean value of Y

**3.8 METHODOLOGY**

**3.8.1 FLOW CHAT SHOWING THE METHODOLOGY**

**Data preparation using Ms Excel**

**Data feeding into Python; Visualization; Feature Selection**

**WEKA API Database**

**Data missing values replacement**

**ZeroR training with Cross- Validation**

**Naïve Bayes Model training with Cross- Validation**

**Cost sensitive J48 Model training with Cross- Validation**

**Get Perfomance results**

**Get Perfomance results**

**Get Perfomance results**

**Compare model performances**

**Design optimal treatment regime for Schizophrenia management**

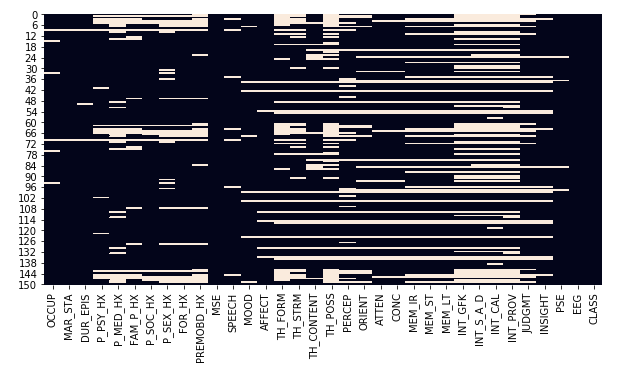
**Figure 4 shows Schematic Diagram of the Methodology**

**3.8.2 DATA PREPARATION, EXPLORATION AND VISUALIZATION**

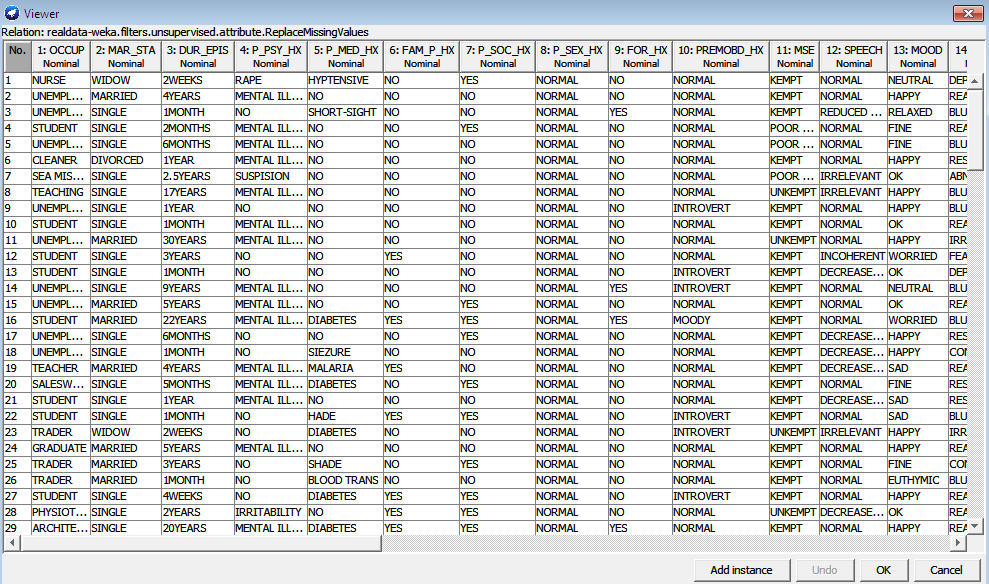
The unstructured file record dataset was converted into electronic structured data using Ms Excel and saved as .csv file acceptable in Waikato Environment for Knowledge Analysis (WEKA) API software. The data was loaded on Python environment for observation and visualization. Four irrelevant features (Y\_O\_REP, AGE, SEX, and DIAGN) were dropped (see Appendix 1 for codes). The description of the cleaned dataset is given in Table 3 below. It is seen clearly that there are missing values in the data as number of entries against some of the features is less than 151.

**Table 3 shows Dataset Description**

|  |
| --- |
| **RangeIndex: 151 entries, 0 to 150** |
| **Data columns (total 34 columns):** |
| **OCCUP 145 non-null object** |
| **MAR\_STA 149 non-null object** |
| **DUR\_EPIS 148 non-null float64** |
| **P\_PSY\_HX 134 non-null object** |
| **P\_MED\_HX 121 non-null object** |
| **FAM\_P\_HX 130 non-null object** |
| **P\_SOC\_HX 138 non-null object** |
| **P\_SEX\_HX 123 non-null object** |
| **FOR\_HX 131 non-null object** |
| **PREMOBD\_HX 123 non-null object** |
| **MSE 151 non-null object** |
| **SPEECH 144 non-null object** |
| **MOOD 143 non-null object** |
| **AFFECT 141 non-null object** |
| **TH\_FORM 108 non-null object** |
| **TH\_STRM 112 non-null object** |
| **TH\_CONTENT 124 non-null object** |
| **TH\_POSS 93 non-null object** |
| **PERCEP 112 non-null object** |
| **ORIENT 120 non-null object** |
| **ATTEN 126 non-null object** |
| **CONC 126 non-null object** |
| **MEM\_IR 116 non-null object** |
| **MEM\_ST 116 non-null object** |
| **MEM\_LT 116 non-null object** |
| **INT\_GFK 89 non-null object** |
| **INT\_S\_A\_D 87 non-null object** |
| **INT\_CAL 84 non-null object** |
| **INT\_PROV 84 non-null object** |
| **JUDGMT 120 non-null object** |
| **INSIGHT 131 non-null object** |
| **PSE 147 non-null object** |
| **EEG 151 non-null object** |
| **CLASS 151 non-null object** |
| **dtypes: float64(1), object(33)** |

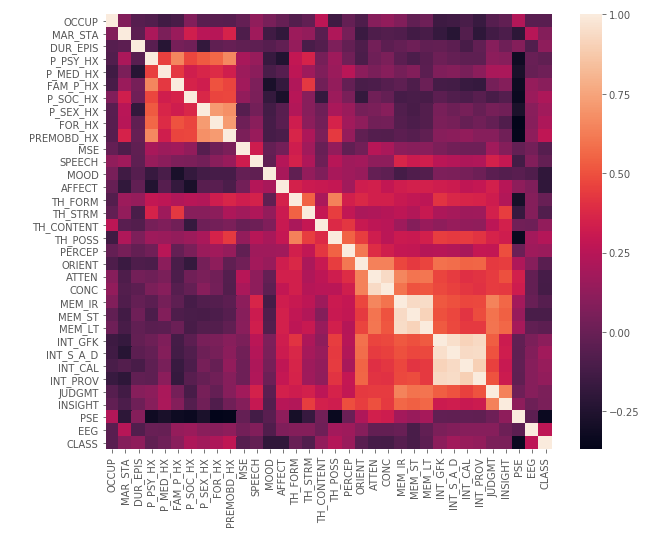
Figure below 5 is a heatmap showing missing value positions in white stripes in dataset. **Figure 5 shows heatmap showing missing value positions in white stripes**

However, missing values were replaced with modal values of corresponding columns using “ReplaceMissingValues” tool in WEKA API (see codes in Appendix 1). The tool replaces nominal missing values with modal values in corresponding columns; the numeric missing values with mean values in corresponding columns.This is necessary in order not to miss a lot of information from the data. Table 4 below shows a section of the data after replacing the missing values.

**Table 4 showing some section of dataset after replacing missing values**

**3.8.3 FEATURE SELECTION**

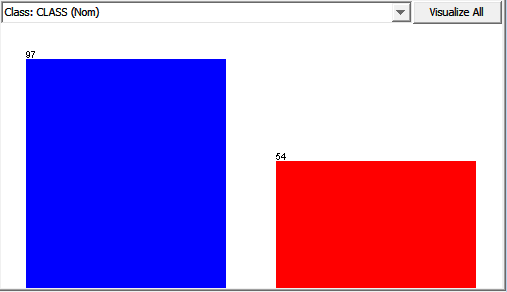
In order to cleanse the dataset of redundant attributes, the dataset was hot-coded and converted to numerical representation. Pearson’s correlation was applied to the data to identify fully correlated features (see Appendix 1 for codes). Pearson’s correlation measures amount of linearity in function defining relation between two attributes. Correlation coefficient between two features has value ranges from -1 through 0 to +1. -1 represent full negative correlation, 0 means no correlation and +1 stands for full positive correlation. It is believed in Data Mining parlance that fully correlated features constitute noise in the dataset. Figure 6 below is a heatmap of correlation values.

** Figure 6 shows Pearson’s Correlation values between each pair of Attributes**

From figure above, it is clear that full correlation exist between ATTEN and CONC, among MEM\_IR, MEM\_ST, MEM\_LT, and among INT\_GFK, INT\_S\_A\_D, INT\_CAL, and INT\_PROV. Consequently 'ATTEN', 'MEM\_ST', 'MEM\_LT', 'INT\_S\_A\_D', 'INT\_CAL', 'INT\_PROV' features were dropped from the dataset to prevent overfitting and for effective model performance.

**3.8.4 CLASS DISTRIBUTION IN DATASET**

Data is visualized in WEKA for evaluation of consistency of the dataset. Figure 7 below shows the stack bar charts of distributions in Class. This distribution shows bias in favour of SCHIZ class.



**Figure 7 shows the stack bar charts of distributions in Class.**

**3.9 MODEL DEVELOPMENT**

Model development involves steps below:

1. DATA RESAMPLING: Data was divided into training and test datasets. The data was divided into 10 folds. The training data was made 90 per cent of the whole dataset, while the test data constitutes remaining 10 per cent, and training was done repeatedly using **Cross-Validation**. (see Apendix 1 for codes)

**Cross validation** is an approach that we can use to estimate the performance of a machine learning algorithm with less variance than a single train-test set split. It works by splitting the dataset into *k*-parts (e.g. *k* = 5 or *k* = 10). Each split of the data is called a fold. The algorithm is trained on *k −* 1 folds with one held back and tested on the held back fold. This is repeated so that each fold of the dataset is given a chance to be the held back test set. After running cross validation you end up with *k* different performance scores that you can summarize using a mean and a standard deviation. The result is a more reliable estimate of the performance of the algorithm on new data. It is more accurate because the algorithm is trained and evaluated multiple times on different data. The choice of *k* must allow the size of each test partition to be large enough to be a reasonable sample of the problem, whilst allowing enough repetitions of the train-test evaluation of the algorithm to provide a fair estimate of the algorithms performance on unseen data. For modest sized datasets in the thousands or tens of thousands of records, *k* values of 3, 5 and 10 are common.

1. MODEL BUILDING: Pruned J48 algorithm was trained using 10-fold cross-validation techniques (see Appendix 3 for model graph). The training data folds were used to train the algorithm. The trained models are shown in the Appendix 2, 4, 5.
2. MODEL EVALUATION: The test data folds were used to evaluate the predictive performance of the model. 131 instances were correctly classified and represent 87 per cent accuracy, while 20 instances representing 13 per cent were wrongly classified. The confusion matrix is shown in the Chapter 4.
3. OPTIMIZATION OF THE MODEL’S PREDICTION: Model’s prediction was optimized by making mistakes costlier than others. Diagnosing a patient negative when actually he is positive can be an expensive mistake. One solution is to minimize the number of False Negatives predictions. J48 algorithm allows us to assign different penalty to different types of errors, in order to discourage a tree from making more costly mistakes. The penalty are designated in a **cost matrix**, which specifies how much costlier each error is relative to any other predictions. It was assumed that False Negative (FN) diagnosis costs the patient two times as much as False Positive (FP) diagnosis. Algorithm awards no cost when an instance is classified correctly, but a FN prediction attracts a penalty cost of 2 versus a False Positive’s cost of 1, as shown in Figure 8 below. The cost-sensitive J48 model produced an accuracy of 78 per cent. This model is shown in the Appendix 2. The Confusion Matrix is shown in Chapter 4.

|  |  |
| --- | --- |
| 0 | 2 |
| 1 | 0 |

**Figure 8 shows cost matrix optimizes prediction performance of J48 Model**

1. BUILDING BASELINER MODEL, ZeroR

The same datasets- training and test-were used to develop the baseline model ZeroR. ZeroR is the most rudiment of classifiers. It simply predicts the majority if data is nominal, and mean value if data is numeric. It over-fits depending on skewness of dataset used. However, it is commonly used to validate the performances of other complex classification algorithms on a given dataset, especially if the dataset is skewed. The model predicted 97 instances accurately and 54 wrongly and as a result had an accuracy of 64 per cent. This is a reflection of skewness in dataset used; the dataset has 97 SCHIZ instances and 54 OTHERS instances.The confusion matrix of ZeroR is shown in chapter 4.

1. BUILDING NAÏVE BAYES MODEL: Naïve Bayes classifier predicts the instance X belongs to the class Ci if and only if:

P(Ci|X) > P(Cj|X) for 1 ≤ j ≤ m, j ≠ I,

where P(Ci|X) =

The Naïve Bayes algorithm was applied to the project dataset and had an accuracy of 82 per cent. When cost matrix was added the accuracy improved to 83 percent. The resulting trained Model template is shown in the Appendix 4.

1. MODEL VALIDATION: ZeroR is a baseline algorithm commonly used in validating machine learning model performance. Naïve Bayes algorithm is a classifier reportedly used by medical scientists in classification problems. Its performances have been adjudged in literatures as reasonable. The two algorithms (ZeroR and Naïve Bayes) were also trained on the project datasets. Performance of cost-sensitive J48 in classifying the instances into appropriate classes was compared with performances of the two validation classifiers to establish the reliance of CS J48 in diagnosis of Schizophrenia.. The comparison of the Baseliner Classifier, ZeroR, ordinary J48, Naïve Bayes, and CS J48 classifiers were done on the basis of Receiver Operating Characteristics (ROC) curve, Areas under ROCs, Confusion Matrices, Sensitivity, Specificity, and Diagnostic Odds Ratio. Results of analysis are structured and presented in Chapter 4.

**3.10 PERFORMANCE STATISTIC METRICS**

**3.10.1 Confusion matrix**

Confusion matrix of a classifier shows the instances of true positives, false negatives, false positives and true negatives in an array. Accuracy is given as

**Accuracy** =

**Table 5 shows a typical Confusion matrix for a classifier**

|  |  |  |
| --- | --- | --- |
| ACTUAL CLASS | PREDICTED CLASS | |
| POSITIVE | NEGATIVE |
| POSITIVE | True positive(TP) | False Negative(FN) |
| NEGATIVE | False Positive(FP) | True Negative(TN) |

**3.10.2 Sensitivity**

It measures the proportion of positives that are correctly identified as such. Perfect

predictor is 100 % sensitive. (Sumathi et al, 2017).

Sensitivity =

**3.10.3 Specificity**

Specificity measures ability of classifier to correctly identify those without Schizophrenia from a dataset. Best specificity is 1.0 whereas worst is 0.0. (Sumathi et al, 2017).

Specificity =

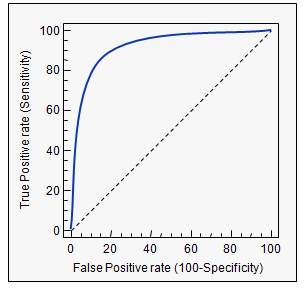
**3.10.4 Diagnostic Odds Ratio**

Diagnostic Odds Ratio, in medical testing, is a measure of effectiveness of a diagnostic test. It is defined as the ratio of the odds of the test being positive if the subject has a disease relative to the odds of the test being positive if the subject does not have the disease. The rational for the Diagnostic Odds Ratio is that it is a single indicator of test performance (like accuracy) but which is independent of prevalence (unlike accuracy) and is presented as an odds ratio, which is familiar to medical practitioners. Mathematically, Diagnostic Odds Ratio, DOR is defined as (Wikipedia, 2018):

Diagnostic Odds Ratio, DOR **= **where TP, FP, FN and TN are the number of true positives, false positives, false negatives, and true negatives respectively. DOR value ranges from zero to infinity, although for useful test it greater than one, and higher diagnostic odds ratios are indicative of better test performance.

**3.10.5 ROC Curves**

In a Receiver Operating Characteristic (ROC) curve the true positive rate (Sensitivity) is plotted in function of the false positive rate (100-Specificity) for different cut-off points. Each point on the ROC curve represents a sensitivity/specificity pair corresponding to a particular decision threshold. A test with perfect discrimination (no overlap in the two distributions) has a ROC curve that passes through the upper left corner (100% sensitivity, 100% specificity) as shown in figure 9 below. Therefore the closer the ROC curve is to the upper left corner, the higher the overall accuracy of the test (Zweig & Campbell, 1993).



**Figure 9 shows ROC curve with perfect discrimination**

**3.10.6 Area under ROC Curve**

Area under ROC Curve (or AUC for short) is a performance metric for binary classification problems. The AUC represents a model’s ability to discriminate between positive and negative classes. An area of 1.0 represents a model that made all predictions perfectly. An area of 0.5 represents a model that is as good as random.

**CHAPTER FOUR**

* 1. **RESULTS AND DISCUSSIONS**

A cost-sensitive J48 (CS J48) pruned tree model was developed using psychometric Health Records dataset acquired from psychiatrists of the Lagos University Teaching Hospital Lagos Nigeria. The model was built using WEKA API imported into Python. Supervised learning regime was implemented. The dataset consists of ninety-seven (97) records of Schizophrenia patients and fifty-four (54) records of patients suffering from other related illnesses, and was defined by 38 features including CLASS column. Irrelevant features not included in DSM-5 criteria for mental health diagnosis were removed. Pearson’s Correlation analysis also revealed that some of the features were data noise and were removed. So the dataset was left with 28 features for model building. Missing values in dataset were replaced using “ReplaceMissingValues” unsupervised filter tool in WEKA API before it was used to train and validate pruned J48 algorithm. This training implemented 10-fold Cross-Validation train and test mode to eliminate bias due to random sampling. Area under ROC curve was 0.943, and with accuracy of 86 percent. The Confusion Matrix is shown in Table 6 below:

**Table 6 Confusion Matrix of pruned J48 Model**

|  |  |  |
| --- | --- | --- |
| ACTUAL CLASS | PREDICTED CLASS | |
| SCHIZ | OTHERS |
| SCHIZ | 81 | 16 |
| OTHERS | 4 | 50 |

The decision mechanism of J48 model was optimized with a cost matrix shown in figure 8. Sc J48 gave accuracy of 78 and ROC area 0.895. The confusion matrix of cost-sensitive J48 model is shown below. One can see that the resulted confusion matrix shows less FN predictions although accuracy was a little bit traded off.

**Table 7 Confusion Matrix of cost-sensitive J48 pruned (CS J48) Model**

|  |  |  |
| --- | --- | --- |
| ACTUAL CLASS | PREDICTED CLASS | |
| SCHIZ | OTHERS |
| SCHIZ | 87 | 10 |
| OTHERS | 23 | 31 |

A ZeroR model was developed using the same dataset and training procedures as above. Area under ROC curve was 0.467 with accuracy of 64 percent. The Confusion Matrix is shown in Table 8 below:

**Table 8 Confusion Matrix of ZeroR Model**

|  |  |  |
| --- | --- | --- |
| ACTUAL CLASS | PREDICTED CLASS | |
| SCHIZ | OTHERS |
| SCHIZ | 97 | 0 |
| OTHERS | 54 | 0 |

A Naïve Bayes model was developed using the same dataset and training procedures as above. Area under ROC curve was 0.917 and with accuracy of 82 percent. The Confusion Matrix is shown in Table 9 below:

**Table 9 Confusion Matrix of Naïve Bayes Model**

|  |  |  |
| --- | --- | --- |
| ACTUAL CLASS | PREDICTED CLASS | |
| SCHIZ | OTHERS |
| SCHIZ | 80 | 17 |
| OTHERS | 10 | 44 |

**4.2 VALIDATION OF MODEL**

The cost-sensitive J48 model developed for quick diagnosis of Schizophrenia was internally validated by Cross-Validation test mode. The external validation was done by comparing its (CS J48’s) performance against performances of J48, Naïve Bayes and ZeroR models developed on the same dataset used in building cost-sensitive J48 model for Schizophrenia diagnosis. The comparison was done on the basis of Area of Receiver Operating Characteristics curve (Area of ROC), accuracy, sensitivity, selectivity, and diagnostic ratio. Tables 10, 11, 12, 13, 14, 15 below show the results of the performance parameters.

**Table 10 shows Accuracies of Models**

|  |  |
| --- | --- |
| MODEL | ACCURACY |
| CS J48 | 78 |
| J48 | 86 |
| NAÏVE BAYES | 82 |
| ZeroR | 64 |

**Table 11 shows Sensitivities of Models**

|  |  |
| --- | --- |
| MODEL | SENSITIVITY |
| CS J48 | 0.897 |
| J48 | 0.835 |
| NAÏVE BAYES | 0.825 |
| ZeroR | 1.000 |

**Table 12 shows Specificities of Models**

|  |  |
| --- | --- |
| MODEL | SPECIFICITY |
| CS J48 | 0.574 |
| J48 | 0.926 |
| NAÏVE BAYES | 0.814 |
| ZeroR | 0 |

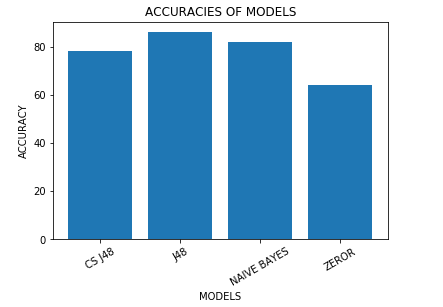
**Table 13 shows Diagnostic Odds Ratios of Models**

|  |  |
| --- | --- |
| MODEL | DOR |
| CS J48 | 21 |
| J48 | 63 |
| NAÏVE BAYES | 21 |
| ZeroR | Undefined |

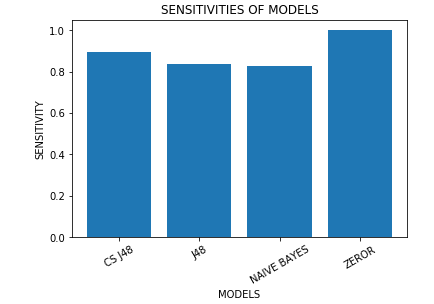
**Table 14 shows Area under Receiver Operating Characteristic Curves of Models**

|  |  |
| --- | --- |
| MODEL | Area Under ROC Curves |
| CS J48 | 0.895 |
| J48 | 0.943 |
| NAÏVE BAYES | 0.917 |
| ZeroR | 0.467 |

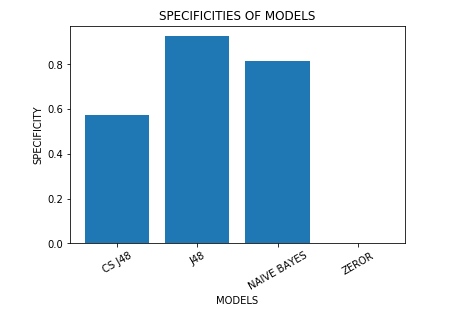
The figures 10 below show Bar Charts comparing different performance metrics for the models.



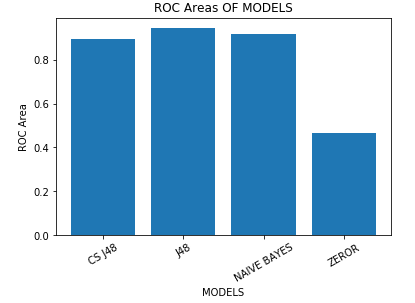
**Figure 10 shows Bar Chart comparing accuracy of each model**

****

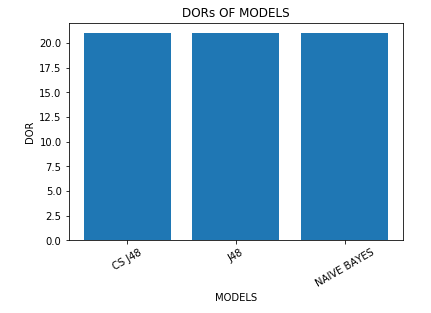
**Figure 11 shows Bar Chart comparing Sensitivity of each model**



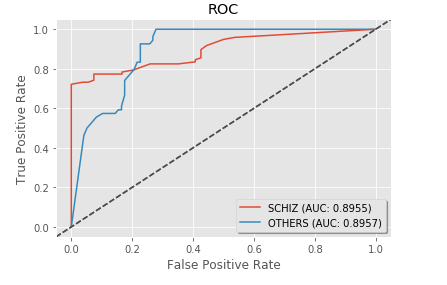
**Figure 12 shows Bar Chart comparing Specificity of each model**



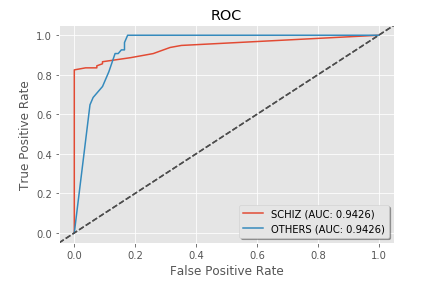
**Figure 13 shows Bar Chart comparing ROC Area of each model**

****

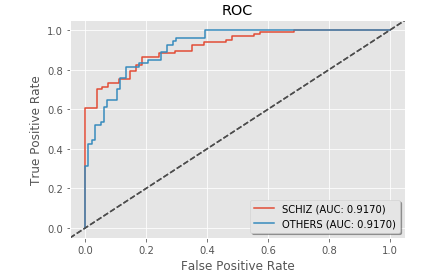
**Figure 14 shows Bar Chart comparing DOR of each model**

****

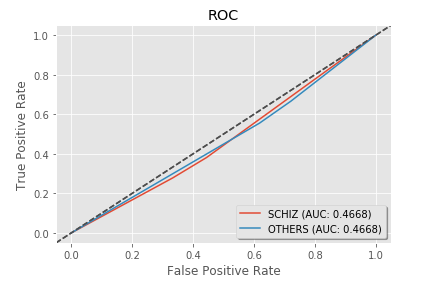
**Figure 15 shows ROC curve of CS J48 model**

****

**Figure 16 shows ROC curve of J48 model**

****

**Figure 17 shows ROC curve of Naïve Bayes model**

****

**Figure 18 shows ROC curve of ZeroR model**

**4.3 DISCUSSIONS**

The results of performance evaluation metrics of the models are presented above. Performance parameters of the cost-sensitive J48 model, J48, ZeroR, and Naïve Bayes are being compared to establish the effectiveness of CS J48 model in Schizophrenia diagnosis. Looking at the performance parameters in figures 10, 11, 12, 13, 14, 15, 16, 17, and 18 above, it is evident that the cost-sensitive model has performed reasonably well as a disease outcome predictor. The CS J48 model’s accuracy is encouraging at 78 per cent and can be reliably used to determine whether a new instance of similar features has Schizophrenia or not. If we look at the sensitivity chart, cost-sensitive J48 is more sensitive than Naïve Bayes and J48. It means its probability of predicting positive instances as positive is higher than that of Naïve Bayes. However both cannot be used to rule in presence of Schizophrenia as their sensitivities are high. CS J48’s specificity is much lower than that of Naive Bayes, this indicates CS J48 is a predictor that can be used to rule out Schizophrenia anytime the outcome is negative. Also, area under Receiver Operating Characteristics (ROC) curve determines the discriminant power of a model. As shown in Figure 13, the areas under the J48’s ROC curve and that of Naïve Bayes’ curve are comparable (0.895 and 0.917 respectively). This means that the two models can discriminate almost equally between positive and negative class instances. The Diagnostic Odd Ratio value is also another performance parameter used in medical field to measure performance of diagnostic tests. Good DOR value is greater than 1 and the greater the better. In this research, CS J48’s DOR, Naïve Bayes’ and J48’s are equal making CS J48 a good diagnostic model. For ZeroR, DOR is undefined indicating that it is a biased predictor. Furthermore, cost-sensitive J48 model minimizes prediction of FN with the cost matrix defined for the algorithm. As shown in the confusion matrices in Tables 6 and 7, the error is more shifted to FP than FN in cost-sensitive J48 model when compared with ordinary J48 model. This assists making diagnosis mistakes that are less costly. Similarly, ROC curves show discriminating power of a model. CS J48 ROC curve shows a clear discrimination of the two classes SCHIZ and OTHERS and with good accuracy, as shown in figure. J48 algorithm was chosen because it is easy to explain and build, and does not require much parameter tuning.

**4.4 PROPOSED OPTIMAL SCHIZOPHRENIA TREATMENT REGIME**

The first-line treatment for patients of Schizophrenia is using medications, antipsychotics: oral and/or Long-Acting Injectable (LAI). The use of LAI is common in patients with history of non-compliance or as choice of preference (David Kenicer et. al., 2016). Antipsychotics are classified as First Generation Antipsychotics (FGAs e.g. haloperidol, chlorpromazine, etc.) and Second Generation Antipsychotics (or Atypical Antipsychotics) (SGAs e.g. clozapine, olanzapine, quetiapine, risperidone, etc.). FGAs act on dopaminergic system by blocking the dopamine type 2 (D2) receptors. This operating mechanism causes extrapyramidal side effects (e.g. dystonia, slurred speech, akathisia, tremor), with some e.g. tardive dyskinesia appearing after long-term exposure to drug (Agency for Healthcare Research and Quality, 2010). Clinical trials have shown that FGAs are effective against positive symptoms of Schizophrenia, and so have been considered ineffective against negative symptoms. SGAs, which are multireceptor antagonists, have been developed to tackle the FGAs’ weakness. They block the brain serotonin and dopamine pathways, and their extrapyramidal side effect profile is favourable. However, they produce other side effect like sedation, sexual dysfunction, hypotension (Agency for Healthcare Research and Quality, 2010).

Determination of optimal treatment regime for individuals affected with Schizophrenia still poses challenges to management of the disease. Individual responses to antipsychotics are not the same. American Psychiatric Association (APA) recommends that psychiatrists should consider both patient’s past responses to the drug and side effect profile in selecting antipsychotics. So selection, dosage, and use duration of medication are subject to observed responses in patient. The process of forcing symptoms to remission should be blended with close monitoring of side effects of drugs, duration, and tolerability for patient. This not-too-easy task makes management of Schizophrenia as daunting as its diagnosis. Many randomized clinical trials have tested general efficacy of antipsychotics or psychosocial intervention, but have failed to present essential steps that could serve as guidance for psychiatrists in daily clinical routines (Stefan Leucht et. al., 2015). Szkultecka-DĊbek et al. (2016) also posited that the optimal treatment regime, in terms of antipsychotic selection, dosage, duration, efficacy and tolerability, for each individual patient is less clear, and patient cooperation is a major concern of all organizations involved in studies of Schizophrenia. In view of these arguments above, to develop an optimal algorithm for treatment and management of Schizophrenia, some important challenges have to be resolved:

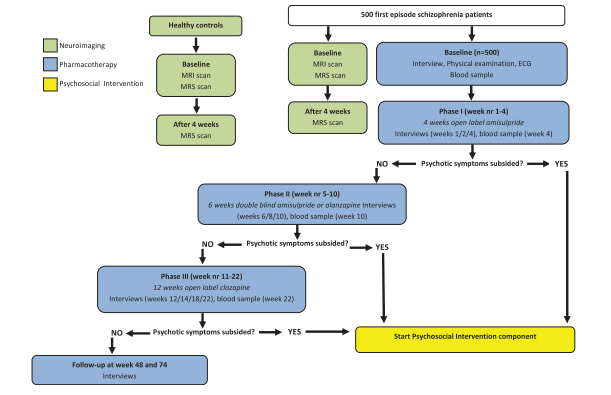
1. If first antipsychotic failed to work, should we increase dose or change drug?
2. At what point should we implement the most potent Food and Drugs Agency (FDA)-approved atypical drug, Clozapine?

Answers to these questions are important, especially in first episode patient, as duration of untreated psychosis is a factor in outcome. There have been attempts to unravel these important questions. Results of those trials indicate no significant difference between efficacy of FGAs and SGAs in tackling symptoms of Schizophrenia. And nor is when drugs are switched. Trials in this category include those conducted by Kinon et al. USA, Klimke et al., Germany, Shalev et al., Israel (Stefan Leucht et al. 2015). However, findings by Cost Utility of Latest Antipsychotic Drugs in Schizophrenia Study (CutLASS) and Clinical Antipsychotic Trials of Intervention Effectiveness (CATIE) reported some differences in effectiveness of FGAs and SGAs in patients with nonrefractory Schizophrenia (Agency for Healthcare Research and Quality, 2010).

In summary, study of prognosis of Schizophrenia is still ongoing. No clear understanding of development course of the disease. Information available on the website of the Institute of Psychiatry, Psychology & Neuroscience (2018), King’s College London reads:

“Despite more than 50 years of pharmacological and psychosocial research, the overall prognosis of Schizophrenia has improved only marginally. This unsatisfactory result may be attributable to three major issues: first, a considerable minority of patients still does not respond sufficiently to current treatments; second, patients who do respond to medication often discontinue it and relapse; finally, even patients who do respond well to treatment and who do adhere to it, continue to suffer from substantial cognitive functional deficits.”

However, European Union’s European First Episode Schizophrenia Trial (EUFEST) presented an algorithm, OPTiMiSE, which could serve a guide for optimal treatment of Schizophrenia. The EUFEST studied the efficacy of switching from one antipsychotic to another in case of non-remission of symptoms, and came up with an algorithm for optimal management of Schizophrenia. The choice of drugs was based on previous efficacy trials conducted by EUFEST. This project adapted the EUFEST algorithm (Figure 19 below) and presents another algorithm (Figure 20 below) that could serve an adjunctive tool for optimal management of Schizophrenia.

****

**Figure 19 shows EUFEST Algorithm for Optimal Treatment of Schizophrenia**

**INPUT SYMPTOMS VECTOR IN WEKA EXPLORER**

**APPLY CS J48 MODEL ON INPUT DATA**

**OTHERS**

**SCHIZ**

**PHASE 1: WEEK 1-4**

**4 WEEKS OPEN LABEL AMISULPRIDE**

**INTERVIEWS (WEEK 1/2/4)**

**BLOOD SAMPLES (WEEK 4)**

**YES**

NO

**PSYCHOTIC SYMPTOMS SUBDUED**

**PHASE 2: WEEK 5-10**

**6 WEEKS DOUBLE BLIND AMISULPRIDE OR OLANZOPINE**

**INTERVIEWS (WEEK 6/8/10)**

**BLOOD SAMPLES (WEEK 10)**

**START PSYCHOSOCIAL INTERVENTION SYSTEM**

**PSYCHOTIC SYMPTOMS SUBDUED**

NO

**YES**

NEW LINE OF TREATMENT

**PSYCHOTIC SYMPTOMS SUBDUED**

**END**

NO

**FOLLOW-UP AT WEEK 48 AND 74**

**INTERVIEWS**

**PHASE 3: WEEK 11-22**

**12 WEEKS OPEN LABEL CLAZOPINE**

**INTERVIEWS (WEEK 12/14/16/18)**

**BLOOD SAMPLES (WEEK 22)**

**YES**

**Figure 20 shows Adapted Algorithm for Optimal Diagnosis and Management of Schizophrenia**

**4.4.1 RATIONAL FOR SELECTING DRUGS IN THE ALGORITHM (UNFEST, 2010, Stefan Leucht et. al., 2015)**

Phase I. The choice of amisulpride as the initial drug was based on the results of the European First Episode Schizophrenia Trial (EUFEST) in which amisulpride turned out as one of the most effective treatments, corroborated by several meta-analyses which suggested a high efficacy and low risk for metabolic and extrapyramidal side-effects. Amisulpride is also a unique “atypical” antipsychotic, because it is a selective D2/D3 (and 5-HT739) receptor antagonist with mesolimbic selectivity rather than a 5-HT2a receptor antagonist.

Phase II. To use olanzapine as the drug to which patients are randomly switched in phase II was again based on the EUFEST trial in which olanzapine together with amisulpride did best in the primary outcome “treatment discontinuation” due to any cause and the proportion of patients in remission applying the criteria defied by Andreasen et  al. Moreover, olanzapine is a multireceptor antagonist and its atypical properties are mainly explained by a stronger antagonism of central serotonin than of central dopamine receptors. It thus has a very  
different receptor binding profile than the selective D2/ D3 receptor antagonist amisulpride. This choice follows the hypothesis that nonresponders to one antipsychotic might respond to another one with a different receptor binding profile for which some evidence is available.

Phase III. Clozapine is considered to be the most efficacious drug for treatment resistant schizophrenia and this has been shown in various individual RCTs and  
meta-analyses, even in nonrefractory patients. Despite this evidence, there currently is on average a 48-month delay before eligible patients are prescribed clozapine in UK.

**CHAPTER FIVE**

**5.1 CONCLUSION**

Quick and accurate detection of Schizophrenia ensures favourable treatment and management outcome as duration of psychotic episode is a factor in efficacy of antipsychotics. This study has designed and developed a cost-sensitive J48 classifier capable of predicting Schizophrenia at an accuracy of 78 per cent. Health records of 151 Schizophrenia patent and those diagnosed otherwise (his& hers) were obtained from the Lagos University Teaching Hospital, Nigeria. The dataset was defined with 38 features including CLASS of instances. The features were got from patient records using DSM-5 Mental Health Diagnosis criteria as guidance. Feature correlation evaluation was done to determine presence of redundant features in dataset. Missing values were replaced using WEKA API tool “ReplaceMissingValue”. The model was trained using supervised learning regime. The psychometric health records data of Schizophrenia patients and those diagnosed of similar illnesses were used to train and validate the classifier using cross-validation procedures. Ten-fold cross-validation technique was used as resampling procedure during model trainng. The prediction of the J48 was optimized using a cost matrix designed to minimize the tendencies of the classifier to predict a patient free of the disease when he/she actually has the disease. Baseline classifier ZeroR and Naïve Bayes classifier were also built using the same dataset. The performance parameters of these classifiers were used for external validation of cost-sensitive J48 model, and measures like Sensitivity, Specificity, Receiver Operating Characteristic (ROC) curves, and Diagnostic Odd Ratio (DOR) were used for analyzing the performances. It was found that the cost-effective J48 model perform considerably well at 78 per cent accuracy against Naïve Bayes (82% accuracy). Naïve Bayes classifier is one of the best classifiers used for medical research as reported in literatures. One of the limitations of the cost-sensitive J48 model is unavailability of sufficiently enough data to train the model. This has effect on generalization power of the model and thus accuracy. The model cannot be used to rule in Schizophrenia completely because of its high sensitivity. However, the model’s overall performance was encouraging.

In this work, an algorithm was also proposed for optimal treatment of first episode Schizophrenia. The algorithm was adapted from the algorithm for optimization of existing treatment for Schizophrenia designed through European Union sponsored European Union First Episode Schizophrenia Trial. The algorithm may serve as a guide to psychiatrists in decision making while design treatment plans for patients, especially first episode patients.

* 1. **LIMITATIONS OF THE CS J48 MODEL**

1. Model was trained using small data size, and hence required retraining with presence of more data.

2. The model has high sensitivity thus could not rule in Schizophrenia completely; however its low specificity makes it appropriate to rule in other diseases.

**5.3 RECOMMENDATIONS**

The following recommendations are suggested for future work on this study:

1. Larger number of records could train the model better in the future for higher accuracy.
2. The model could be developed into mobile application for public use
3. Other types of feature selection algorithms could be tried to observe performance
4. Other missing values treatment procedure could be explored in future.
5. A critical research into unravel prognosis of Schizophrenia is a good future feast.
6. Other data mining algorithms could be used in future, especially Deep Learning procedures.

**APPENDIX 1**

**Sequence of Codes for the Project Work in Jupiter Notebook**

import os

import seaborn as sn

import pandas as pd

from pandas import read\_csv

project\_data = pd.read\_csv('used\_PROJECT\_DATANEW.csv')

project\_data.head()

#Drop irrelevant columns an save remaiming dataset as data

project\_data.drop(['Y\_O\_REP','AGE','SEX','DIAGN'], axis = 1, inplace = True)

project\_data.to\_csv('cleaned\_data.csv', index=False)

project\_data.head(5)

project\_data.info() #Data exploration

import matplotlib.pyplot as plt

plt.figure(figsize=(10,5))

sn.heatmap(project\_data.isnull(),cbar=False) #heatmap showing missing values

# Handling Categorical values and perform correlation for feature selection

import matplotlib.pyplot as plt

from matplotlib import style

style.use('ggplot')

import numpy as np

import pandas as pd

#project\_data.drop(['Y\_O\_REP','AGE','SEX','DIAGN'], axis = 1, inplace = True)

project\_data.convert\_objects(convert\_numeric=True)

#data.fillna(0, inplace=True)

def handle\_non\_num\_data(data):

columns=data.columns.values

for column in columns:

text\_digit\_vals={}

def convert\_to\_int(val):

return text\_digit\_vals[val]

if data[column].dtype != np.int64 and data[column].dtype != np.float64:

column\_contents=data[column].values.tolist()

unique\_elements=set(column\_contents)

x = 0

for unique in unique\_elements:

if unique not in text\_digit\_vals:

text\_digit\_vals[unique] = x

x+=1

data[column]=list(map(convert\_to\_int, data[column]))

return data

project\_data= handle\_non\_num\_data(project\_data)

#print(project\_data.head())

#print(project\_data.CLASS)

plt.figure(figsize=(10,8))

sn.heatmap(project\_data.corr(),cbar=True) #heatmpa showing correlation between features

cleaned\_data=pd.read\_csv('cleaned\_data.csv') #data cleaning of redundants

cleaned\_data.drop(['ATTEN','MEM\_ST','MEM\_LT','INT\_S\_A\_D','INT\_CAL','INT\_PROV'], axis = 1, inplace = True)

cleaned\_data.to\_csv('cleaned\_data.csv', index=False) #save cleaned data to directory

cleaned\_data.head(3)

import weka.core.jvm as jvm #import python-weka wrapper

jvm.start(packages=True) # starting java virtual machine environment for WEKA packages

import weka.core.converters as conv #import data to WEKA API

cleaned\_data=conv.load\_any\_file('cleaned\_data.csv')

cleaned\_data.class\_is\_last()

#COST-SENSITIVE J48 CLASSIFIER

from weka.filters import Filter

replace = Filter(classname="weka.filters.unsupervised.attribute.ReplaceMissingValues")#Replace missing values

from weka.classifiers import Classifier,SingleClassifierEnhancer

cls = SingleClassifierEnhancer(classname="weka.classifiers.meta.CostSensitiveClassifier", options=["-cost-matrix","[0 2;1 0]","-S","2"])

base = Classifier(classname="weka.classifiers.trees.J48", options=["-C", "0.25"])

cls.classifier=base

from weka.classifiers import FilteredClassifier,PredictionOutput

fc = FilteredClassifier()

fc.filter = replace

fc.classifier = cls

pout = PredictionOutput(classname="weka.classifiers.evaluation.output.prediction.PlainText")

from weka.classifiers import Evaluation

from weka.core.classes import Random

evl = Evaluation(cleaned\_data)

evl.crossvalidate\_model(fc, cleaned\_data, 10, Random(1))

print(evl.percent\_correct)

print(evl.summary())

print(evl.class\_details())

print("=====Confusion matrix=====")

print(evl.confusion\_matrix)

#print(pout.buffer\_content())

import weka.plot.classifiers as plcls # NB: matplotlib is required

plcls.plot\_roc(evl, class\_index=[0, 1.5], wait=True)

cls.build\_classifier(cleaned\_data)

#for index, inst in enumerate(cleaned\_data):

#pred = cls.classify\_instance(inst)

#dist=cls.distribution\_for\_instance(inst)

#print(str(index+1)+ ": Label Index =" + str(pred))

print(cls)

#import weka.plot.graph as graph # NB: pygraphviz and PIL are required

#graph.plot\_dot\_graph(cls.graph)

#NAIVE BAYES CLASSIFIER

cls = Classifier(classname="weka.classifiers.bayes.NaiveBayes")

from weka.classifiers import FilteredClassifier,PredictionOutput

fc = FilteredClassifier()

fc.filter = replace

fc.classifier = cls

pout = PredictionOutput(classname="weka.classifiers.evaluation.output.prediction.PlainText")

from weka.classifiers import Evaluation

from weka.core.classes import Random

evl = Evaluation(cleaned\_data)

evl.crossvalidate\_model(fc, cleaned\_data, 10, Random(1))

print(evl.percent\_correct)

print(evl.summary())

print(evl.class\_details())

print(evl.confusion\_matrix)

print(pout.buffer\_content())

#print("confusionMatrix: " + str(eval.confusion\_matrix))

#print("areaUnderROC/1: " + str(eval.area\_under\_roc(1),class\_index=[0, 1]))

cls.build\_classifier(cleaned\_data)

#for index, inst in enumerate(cleaned\_data):

#pred = cls.classify\_instance(inst)

#print(str(index+1)+ ": Label Index =" + str(pred))

import weka.plot.classifiers as plcls # NB: matplotlib is required

plcls.plot\_roc(evl, class\_index=[0, 1], wait=True)

#import weka.plot.graph as graph # NB: pygraphviz and PIL are required

#graph.plot\_dot\_graph(cls.graph)

print(cls)

#BASELINER CLASSIFIER ZeroR

cls = Classifier(classname="weka.classifiers.rules.ZeroR")

from weka.classifiers import FilteredClassifier,PredictionOutput

fc = FilteredClassifier()

fc.filter = replace

fc.classifier = cls

pout = PredictionOutput(classname="weka.classifiers.evaluation.output.prediction.PlainText")

from weka.classifiers import Evaluation

from weka.core.classes import Random

evl = Evaluation(cleaned\_data)

evl.crossvalidate\_model(fc, cleaned\_data, 10, Random(1))

print(evl.percent\_correct)

print(evl.summary())

print(evl.class\_details())

print(evl.confusion\_matrix)

print(pout.buffer\_content())

#print("confusionMatrix: " + str(eval.confusion\_matrix))

#print("areaUnderROC/1: " + str(eval.area\_under\_roc(1),class\_index=[0, 1]))

#for index, inst in enumerate(cleaned\_data):

#pred = cls.classify\_instance(inst)

#print(str(index+1)+ ": Label Index =" + str(pred))

cls.build\_classifier(cleaned\_data)

print(cls)

import weka.plot.classifiers as plcls # NB: matplotlib is required

plcls.plot\_roc(evl, class\_index=[0, 1], wait=True)

#ORDINARY J48 CLASSIFIER

cls = Classifier(classname="weka.classifiers.trees.J48", options=["-C", "0.25"])

from weka.classifiers import FilteredClassifier,PredictionOutput

fc = FilteredClassifier()

fc.filter = replace

fc.classifier = cls

pout = PredictionOutput(classname="weka.classifiers.evaluation.output.prediction.PlainText")

from weka.classifiers import Evaluation

from weka.core.classes import Random

evl = Evaluation(cleaned\_data)

evl.crossvalidate\_model(fc, cleaned\_data, 10, Random(1))

print(evl.percent\_correct)

print(evl.summary())

print(evl.class\_details())

print(evl.confusion\_matrix)

print(pout.buffer\_content())

#print("confusionMatrix: " + str(eval.confusion\_matrix))

#print("areaUnderROC/1: " + str(eval.area\_under\_roc(1),class\_index=[0, 1]))

#for index, inst in enumerate(cleaned\_data):

#pred = classify\_instance(inst)

#print(str(index+1)+ ": Label Index =" + str(pred))

import weka.plot.classifiers as plcls # NB: matplotlib is required

plcls.plot\_roc(evl, class\_index=[0, 1], wait=True)

#import weka.plot.graph as graph # NB: pygraphviz and PIL are required

#graph.plot\_dot\_graph(cls.graph)

cls.build\_classifier(cleaned\_data)

print(cls)

**APPENDIX 2**

**CS J48 Model**

CostSensitiveClassifier using reweighted training instances

weka.classifiers.trees.J48 -C 0.25 -M 2

Classifier Model

J48 pruned tree

------------------

PSE = GOOD: SCHIZ (59.14)

PSE = NORMAL

| PREMOBD\_HX = NORMAL

| | P\_MED\_HX = HYPTENSIVE: SCHIZ (0.0)

| | P\_MED\_HX = SHORT-SIGHT: SCHIZ (0.0)

| | P\_MED\_HX = NO

| | | MEM\_IR = GOOD: SCHIZ (22.07/3.37)

| | | MEM\_IR = FAIR: SCHIZ (3.33/0.1)

| | | MEM\_IR = POOR: OTHERS (5.0/1.2)

| | P\_MED\_HX = DIABETES: SCHIZ (5.18/0.3)

| | P\_MED\_HX = SIEZURE: OTHERS (1.94)

| | P\_MED\_HX = MALARIA: SCHIZ (0.0)

| | P\_MED\_HX = HADE: SCHIZ (0.0)

| | P\_MED\_HX = SHADE

| | | CONC = GOOD: OTHERS (6.76/0.75)

| | | CONC = POOR: SCHIZ (4.25/0.13)

| | | CONC = REDUCED: OTHERS (0.0)

| | P\_MED\_HX = BLOOD TRANS: SCHIZ (2.59/0.15)

| | P\_MED\_HX = ALLEGIES: SCHIZ (0.0)

| | P\_MED\_HX = HbSS: OTHERS (3.25/1.23)

| | P\_MED\_HX = STROKE: SCHIZ (0.0)

| | P\_MED\_HX = HDASP: OTHERS (1.94)

| | P\_MED\_HX = IMMUNE DISORDER: SCHIZ (0.0)

| | P\_MED\_HX = ASPHYXIA: OTHERS (1.94)

| | P\_MED\_HX = JAUNDICE: OTHERS (1.94)

| | P\_MED\_HX = GLAUCOMA: SCHIZ (0.0)

| | P\_MED\_HX = ASTHMA: OTHERS (1.94)

| PREMOBD\_HX = INTROVERT: SCHIZ (9.74)

| PREMOBD\_HX = MOODY: SCHIZ (0.0)

| PREMOBD\_HX = ASTHMA: SCHIZ (2.44)

| PREMOBD\_HX = MELANCHOLIC: OTHERS (1.83)

PSE = PALE

| INT\_GFK = FAIR: SCHIZ (4.49/0.52)

| INT\_GFK = GOOD: OTHERS (6.74/2.31)

| INT\_GFK = POOR: SCHIZ (4.49/0.52)

Number of Leaves : 30

Size of the tree : 36

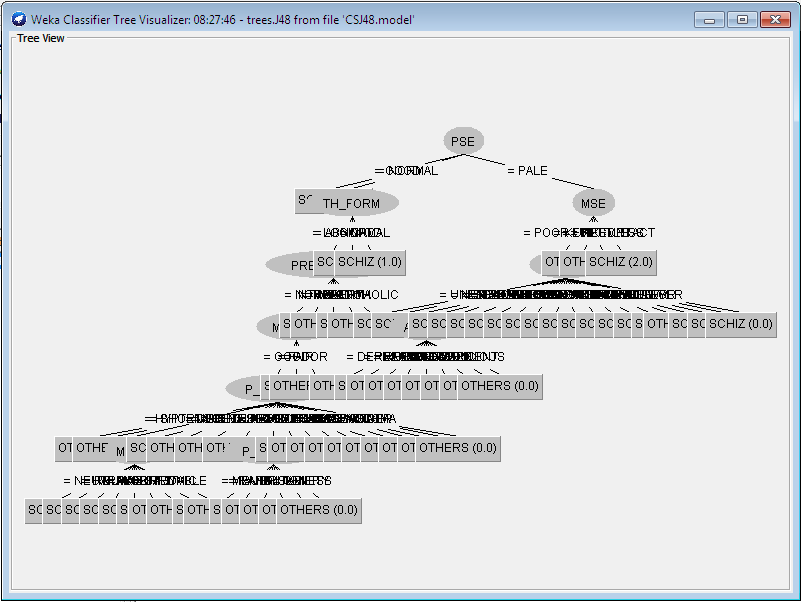
Cost Matrix

0 2

1 0

**APPENDIX 3**

**Cost-sensitive J48 Model tree**



**Figure 21 shows Cost-sensitive J48 Model tree**

**APPENDIX 4**

**Naïve Bayes Model**

Naive Bayes Classifier

Class

Attribute SCHIZ OTHERS

(0.64) (0.36)

=============================================

OCCUP

NURSE 4.0 1.0

UNEMPLOYED 29.0 10.0

STUDENT 27.0 22.0

CLEANER 4.0 1.0

SEA MISTRESS 4.0 1.0

TEACHING 4.0 1.0

TEACHER 3.0 1.0

SALESWOMAN 3.0 1.0

TRADER 9.0 10.0

GRADUATE 3.0 1.0

PHYSIOTHERAPIST 3.0 1.0

ARCHITECT 3.0 1.0

GUARD 3.0 1.0

ASSISTANT 3.0 1.0

RETIRED 4.0 1.0

ENGINEER 1.0 4.0

WRITER 1.0 4.0

COBBLER 1.0 4.0

AUDITOR 1.0 4.0

LAWYER 1.0 4.0

[total] 111.0 74.0

MAR\_STA

WIDOW 9.0 1.0

MARRIED 24.0 16.0

SINGLE 62.0 37.0

DIVORCED 4.0 4.0

[total] 99.0 58.0

DUR\_EPIS

mean 51.2503 75.5045

std. dev. 80.8657115.2294

weight sum 95 53

precision 16.6739 16.6739

P\_PSY\_HX

RAPE 4.0 1.0

NO 35.0 10.0

SUSPISION 4.0 1.0

MENTAL ILLNESS 42.0 40.0

IRRITABILITY 3.0 1.0

GRIEF 1.0 4.0

[total] 89.0 57.0

P\_MED\_HX

HYPTENSIVE 4.0 1.0

SHORT-SIGHT 4.0 1.0

NO 35.0 13.0

DIABETES 11.0 1.0

SIEZURE 3.0 4.0

MALARIA 3.0 1.0

HADE 3.0 1.0

SHADE 6.0 10.0

BLOOD TRANS 3.0 1.0

ALLEGIES 3.0 1.0

HbSS 5.0 4.0

STROKE 3.0 4.0

HDASP 3.0 4.0

IMMUNE DISORDER 1.0 4.0

ASPHYXIA 1.0 4.0

JAUNDICE 1.0 4.0

GLAUCOMA 1.0 4.0

ASTHMA 1.0 4.0

[total] 91.0 66.0

FAM\_P\_HX

NO 61.0 36.0

YES 20.0 17.0

[total] 81.0 53.0

P\_SOC\_HX

YES 34.0 18.0

NO 52.0 38.0

[total] 86.0 56.0

P\_SEX\_HX

NORMAL 73.0 46.0

MASTURBATE 1.0 4.0

GORNORREA 1.0 4.0

[total] 75.0 54.0

FOR\_HX

NO 72.0 51.0

YES 10.0 2.0

[total] 82.0 53.0

PREMOBD\_HX

NORMAL 54.0 49.0

INTROVERT 16.0 1.0

MOODY 3.0 1.0

ASTHMA 3.0 1.0

MELANCHOLIC 1.0 4.0

[total] 77.0 56.0

MSE

KEMPT 75.0 42.0

POOR EYE CONTACT 13.0 7.0

UNKEMPT 10.0 5.0

RESTLESS 3.0 4.0

[total] 101.0 58.0

SPEECH

NORMAL 62.0 38.0

REDUCED VOL 4.0 1.0

IRRELEVANT 9.0 1.0

INCOHERENT 3.0 1.0

DECREASED TONE 15.0 7.0

MUTE 3.0 6.0

SLURRED 1.0 4.0

INCREASED TONE 1.0 4.0

[total] 98.0 62.0

MOOD

NEUTRAL 8.0 7.0

HAPPY 32.0 19.0

RELAXED 4.0 1.0

FINE 14.0 7.0

OK 21.0 7.0

WORRIED 5.0 1.0

SAD 9.0 10.0

EUTHYMIC 5.0 7.0

IRRITABLE 3.0 1.0

[total] 101.0 60.0

AFFECT

DEPRESSED 6.0 7.0

REACTIVE 35.0 22.0

BLUNT 23.0 4.0

RESTRICTED 18.0 4.0

ABNORMAL 4.0 2.0

IRRITABLE 5.0 4.0

FEARFUL 3.0 1.0

CONGRUENT 7.0 10.0

SUSPICIOUS 3.0 1.0

[total] 104.0 55.0

TH\_FORM

LOGICAL 46.0 43.0

ABNORMAL 21.0 1.0

NFTD 2.0 1.0

[total] 69.0 45.0

TH\_STRM

REDUCED 18.0 1.0

NO 4.0 1.0

NORMAL 45.0 37.0

INCREASED 7.0 7.0

[total] 74.0 46.0

TH\_CONTENT

PERSECUTORY DELUSION 40.0 13.0

NORMAL 25.0 22.0

OBSESSION 6.0 4.0

AUDITORY HALLUCINATION 4.0 1.0

DELUSION 6.0 1.0

DELUSION OF REFERENCE 3.0 1.0

GRANDIOSE DELUSION 3.0 1.0

AUDI TORY DELUSION 3.0 1.0

DISORDER 1.0 4.0

GRANDEUR DELUSION 1.0 4.0

[total] 92.0 52.0

TH\_POSS

IMPAIRED 8.0 7.0

NORMAL 45.0 37.0

[total] 53.0 44.0

PERCEP

AUDITORY HALLUCINATION 35.0 7.0

FUNCTIONAL HALLUCINATION 4.0 1.0

SOMATIC HALLUCINATION 6.0 1.0

TACTILE HALLUCINATION 5.0 4.0

OLFACTOTY HALLUCINATION 3.0 1.0

HALLUCINATION 7.0 1.0

VISUAL HALLUCINATION 7.0 4.0

PREOCCUPATION 3.0 1.0

NO 12.0 28.0

[total] 82.0 48.0

ORIENT

ORIENTED IN TPP 77.0 43.0

IMPAIRED 3.0 1.0

[total] 80.0 44.0

CONC

GOOD 63.0 37.0

POOR 16.0 7.0

REDUCED 8.0 1.0

[total] 87.0 45.0

MEM\_IR

GOOD 68.0 37.0

FAIR 6.0 1.0

POOR 3.0 7.0

[total] 77.0 45.0

INT\_GFK

FAIR 12.0 1.0

GOOD 34.0 34.0

POOR 7.0 7.0

[total] 53.0 42.0

JUDGMT

POOR 42.0 25.0

GOOD 35.0 22.0

[total] 77.0 47.0

INSIGHT

PARTIAL 27.0 22.0

POOR 41.0 13.0

PERSISTENT 4.0 1.0

GOOD 18.0 13.0

[total] 90.0 49.0

PSE

GOOD 48.0 1.0

NORMAL 39.0 46.0

PALE 9.0 10.0

[total] 96.0 57.0

EEG

NORMAL 98.0 49.0

ALTERED 1.0 7.0

[total] 99.0 56.0

**APPENDIX 5**

**J48 Model**

J48 pruned tree

------------------

PSE = GOOD: SCHIZ (48.28)

PSE = NORMAL

| PREMOBD\_HX = NORMAL

| | TH\_FORM = LOGICAL

| | | P\_MED\_HX = HYPTENSIVE: OTHERS (0.0)

| | | P\_MED\_HX = SHORT-SIGHT: OTHERS (0.0)

| | | P\_MED\_HX = NO

| | | | CONC = GOOD

| | | | | TH\_POSS = IMPAIRED: OTHERS (3.13)

| | | | | TH\_POSS = NORMAL: SCHIZ (13.72/1.72)

| | | | CONC = POOR: OTHERS (6.26)

| | | | CONC = REDUCED: SCHIZ (0.0)

| | | P\_MED\_HX = DIABETES: SCHIZ (2.2/0.2)

| | | P\_MED\_HX = SIEZURE: OTHERS (3.3)

| | | P\_MED\_HX = MALARIA: OTHERS (0.0)

| | | P\_MED\_HX = HADE: OTHERS (0.0)

| | | P\_MED\_HX = SHADE

| | | | CONC = GOOD: OTHERS (11.04/0.86)

| | | | CONC = POOR: SCHIZ (3.41/0.14)

| | | | CONC = REDUCED: OTHERS (0.0)

| | | P\_MED\_HX = BLOOD TRANS: SCHIZ (1.84/0.17)

| | | P\_MED\_HX = ALLEGIES: OTHERS (0.0)

| | | P\_MED\_HX = HbSS: OTHERS (3.3)

| | | P\_MED\_HX = STROKE: OTHERS (0.0)

| | | P\_MED\_HX = HDASP: OTHERS (3.3)

| | | P\_MED\_HX = IMMUNE DISORDER: OTHERS (0.0)

| | | P\_MED\_HX = ASPHYXIA: OTHERS (2.76)

| | | P\_MED\_HX = JAUNDICE: OTHERS (2.76)

| | | P\_MED\_HX = GLAUCOMA: OTHERS (0.0)

| | | P\_MED\_HX = ASTHMA: OTHERS (3.3)

| | TH\_FORM = ABNORMAL: SCHIZ (10.77/1.34)

| | TH\_FORM = NFTD: SCHIZ (1.18/0.15)

| PREMOBD\_HX = INTROVERT: SCHIZ (8.0)

| PREMOBD\_HX = MOODY: OTHERS (0.0)

| PREMOBD\_HX = ASTHMA: SCHIZ (2.0)

| PREMOBD\_HX = MELANCHOLIC: OTHERS (3.0)

PSE = PALE

| INT\_GFK = FAIR: SCHIZ (3.49/0.6)

| INT\_GFK = GOOD: OTHERS (10.48/2.68)

| INT\_GFK = POOR: SCHIZ (3.49/0.6)

Number of Leaves : 33

Size of the tree : 41

**APPENDIX 6**

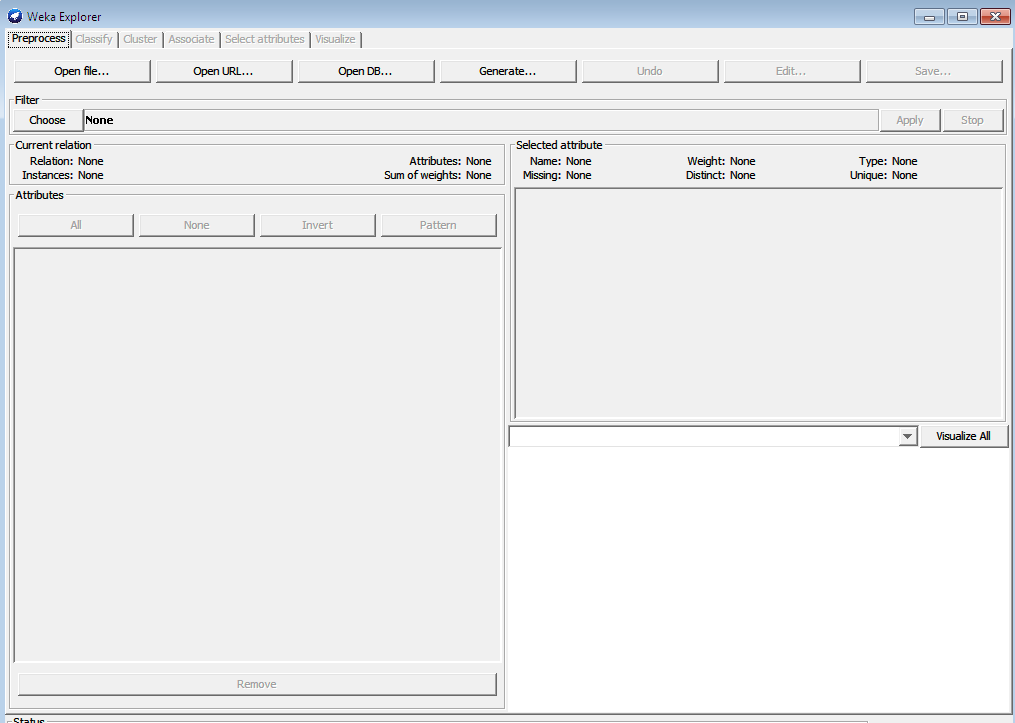
**Guide To Use the Model by Any Psychiatrist for Schizophrenia Diagnosis**

**STEP 1:** Prepare the input vector of patient symptom features in Ms. Excel under the headings below and save as csv file. The format is shown below using Table 1 as guidance. Leave Class column empty.

**Table 15 shows Feature Vector Sample**

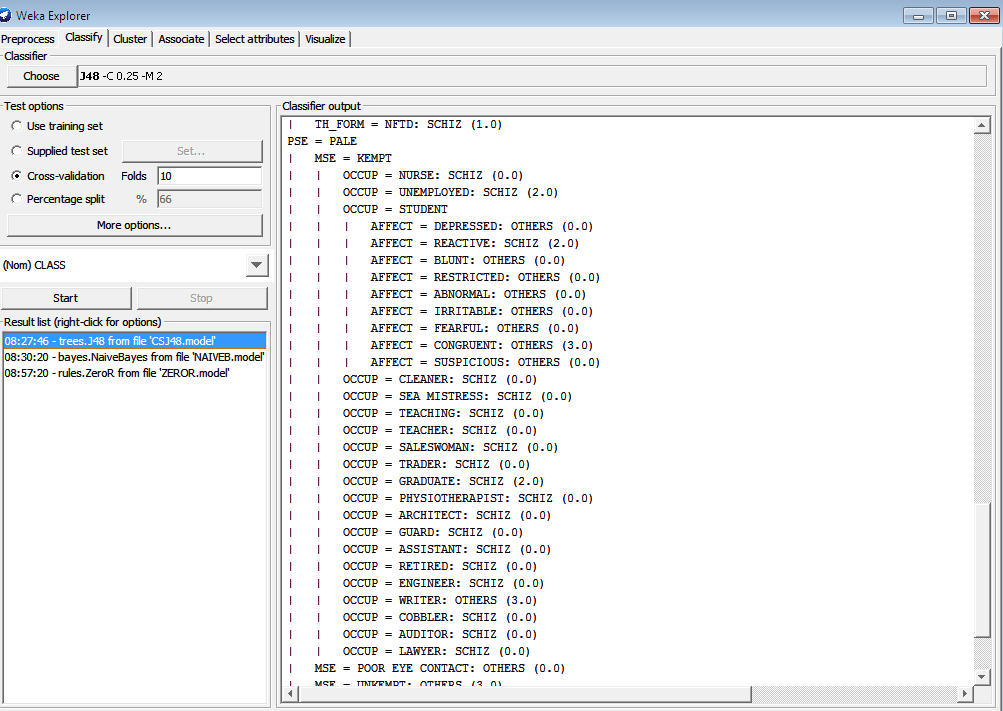


**STEP 2:** Open the file in WEKA Explorer (WEKA is free open software), remove redundant features, and click on Classify button. See figure below.



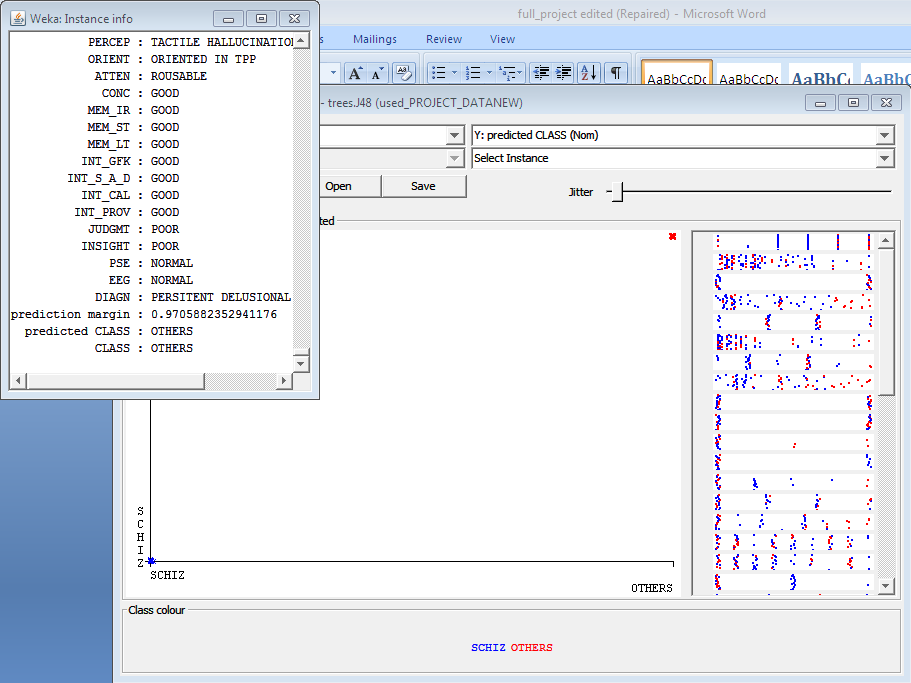
**Figure 22 shows WEKA Explorer Environment**

**STEP 3:** Right-click the object pane of Explorer and load the CS J48 model; right-click on the loaded model and select re-apply model and then click start button. See figure 18 below



**Figure 23 shows WEKA Object Pane**

**STEP 4:** Right-click on the new model formed on clicking start button and select visualize classifier error option. Move the Jitter slide to see plotted point on Class Vs. Predicted Class plot; click on the point and get the result of classification. See figure 24 below.



**Figure 24 shows WEKA Classifier Errors Pane**

**STEP 5:** Follow the adapted Optimization Algorithm in Figure 20 for optimized management course.

**REFERENCES**

MsSumathi M.R., Poorna B. (2016). Prediction of Mental Health Problems Among  
Children Using Machine Learning Techniques.International Journal of Advanced Computer Science and Applications, 7(1).

Kipli, K., Abbas Z.K., and Isredza, R.A.H. (2013). Investigating Machine Learning Techniques for Detection of Depression using Structural MRI Volumetric Features. International Journal of Bioscience, Biochemistry and Bioinformatics 3(5), 444-448.

Masri, R.Y., Jani, H.M. (2012). Employing Artificial Intelligence Techniques in  
Mental Health Diagnostic Expert. System.Computer & InformationScience (ICCIS) International Conference 1, 495-499.

Babita P., Depika K. (2017). Diagnosis of EEG-based Diseases using Data Mining and Case-Based Reasoning. International Journal of Intelligent Home, 1(1-2)

Sarina J. et at (2013). Classification of Schizophrenia using MR Imaging Data.Retrieved from http://www.frontialsin .org/articles/10.3389/fpsyt.2013.00095/full

Stefan P. Koch et al (2015). Diagnostic Classification of Schizophrenia Patients on the Basis of Regional Reward-related fMRI Signal Patterns. Retrieved from http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0119089

Krishnaveni et al.(2016). Diagnosis and Evaluation of ADHD using Naïve Bayes and J48 Classifiers. Retrieved from htp://ncbi.nlm.nih.gov/pmc/article/PMC4370557/Retrievedfrom<http://ieeexplore.ieee.org/document/7724579>

Miseon S. et al. (2016). Machine-Learning-Based Diagnosis of Schizophrenia using Combined Sensor-level and Source-level EEG Levels. RESEARCH, 176( 2-3), 314-317

Shaoqiang H. et al (2017). Recognition of Early-Onset Schizophrenia using Deep-Learning Method. Retrieved from <http://link.springer.com/article/10.1186/s40535-017-0044-3>

Filip, B. and Jesus, J.C.(2015).A Neural Network Based Model for Predicting Psychological Conditions. Retrieved from http://link.springer.com/chapter/10.1007%2F978-3-319-23344-4\_25

Sumathi M.R., Poorna B.(2017).Design and Development of Ensemble of Naïve Bayes Classifiers to Predict Social and Communication Deficiency Among Children. International Journal of Applied Engineering Research, 12(24), 14190-14198.

Li, Y., Yang, L., Wang, P., Zhang, C., Xiao, J., Zhang, Y., &Qiu, M. (2017). Classification of Parkinson's Disease by Decision Tree Based Instance Selection and Ensemble Learning Algorithms. Journal of MedicalImaging and Health Informatics, 7(2), 444-452.

Lee, E. S. (2017, August). Exploring the Performance of Stacking Classifier to Predict Depression Among the Elderly. Healthcare Informatics (ICHI), 2017 IEEEInternational Conference on (pp. 13-20). IEEE.

Husain, W., Xin, L. K., &Jothi, N. (2016, August). Predicting Generalized Anxiety Disorder Among Women Using Random Forest Approach. In Computer and Information Sciences (ICCOINS), 2016 3rdInternational Conference on (pp. 37- 42). IEEE.

Abou-Warda H., Belal N.A., El-Sonbaty Y., Darwish S. (2017). A Random Forest Model for Mental Disorders Diagnostic Systems. In: Hassanien A., Shaalan K., Gaber T., Azar A., Tolba M. (eds.) Proceedings of the International Conference on Advanced Intelligent Systems and Informatics AISI 2016. Advances in Intelligent Systems and Computing, Vol. 533. Springer, Cham

Jaree T., Vatinee S., Verayut, M. (2016). Predicting Schizophrenia at Risk of Readmissions in the Short-and-Long-Term using Decision Tree Model.KKU Research Journal,21(3), 91-103.

Monika S.D. et. al. (2016).Treatment Patterns of Schizophrenia Based on The Data From Seven Central and Eastern European Countries. Psychiatria Danubina, 28(3), 234-242

Reetu, Narender K.(2015).Medical Diagnosis for Liver Cancer using Classification Techniques. International Journal of Recent Scientific Research, 6(6), 4809-4813.

Vijayarani, S., Dhayanand, S. (2015). Liver Disease Prediction using SVM and Naive Bayes Algorithms. International Journal of Science, Engineering and Technology Research (IJSETR), 4(4), 816- 820.

Neelamegam, S., Ramaraj, E., (2013). Classification Algorithm in Data Mining: An Overview. International Journal of P2P Network Trends and Technology (IJPTT), 4(8), 369-374.

Rajiv T. et al. (2013). Definition of Description of Schizophrenia in the DSM-V.Schizophr. Res. Retrieved from <http://dx.doi.org/10.1016/j.schres.2013.05.028>

Serlin Merdan (2018). Optimization and Machine Learning Methods for Diagnosis of Prostate Cancer. Retrieved from <http://deepblue.lib.umich.edu/handle/2027.42/143976>.

Amresh Srivastava(2009). Switching and Selecting Atypical Antipsychotic Drugs: Quetiapine. Retrieved from <http://ir.lib.uwo.ca/psychiatrypres>

Stefan Leucht et. al. (2015). The Optimization of Treatment and Management of Schizophrenia in Europe (OPTiMiSE) Trial: Rationale for Its Methodology and a Review of Effectiveness of Switching Antipsychotic…Retrieved from :https://www.researchgate.net/publication/273783580

Shalev A, Hermesh H, Rothberg J, Munitz H. Poor Neuroleptic Response in Acutely Exacerbated Schizophrenic Patients. Acta Psychiatr Scand. 1993;87:86–91.

Klimke A, Klieser E, Lehmann E, Miele L. Initial Improvement as a Criterion for Drug Choice in Acute Schizophrenia. Pharmacopsychiatry. 1993; 26:25–29.

TahaSamadSoltaniHeris, MostafaLangarizadeh, Zahra Mahmoodvand, Maryam Zolnoori Intelligent Diagnosis of Asthma Using Machine Learning Algorithms. International Research Journal of Applied and Basic Sciences. ISSN 2251 838X / Vol, 4 (12) (2013) 4041 -4046.

Swapna G., Vinayakumar R., Soman K.P. Diabetes detection using deep learning algorithms. ICT Express 4 (2018) 243–246.

Abraham Karplus. Machine Learning Algorithms for Cancer Diagnosis. Santa Cruz County Science Fair 2012.

Dinu A.J., Ganesan R1, Felix Joseph2 and Balaji V. A study on Deep Machine Learning Algorithms for diagnosis of diseases. International Journal of Applied Engineering Research ISSN 0973-4562 Volume 12, Number 17 (2017) 6338-6346. Research India Publications. Retrieved from <http://www.ripublication.com>

Shashikant U. Ghumbre and Ashok A. Ghatol. Heart Disease Diagnosis Using Machine Learning Algorithm (2012).

Fahad Saeed. Towards quantifying psychiatric diagnosis using machine learning algorithms and big fMRI data. Saeed Big Data Analytics (2018) <https://doi.org/10.1186/s41044-018-0033-0>

Lei Chu, Robert Qiu, Haichun Liu, Zenan Ling, Tianhong Zhang and Jijun Wang. Individual Recognition in Schizophrenia using Deep Learning Methods with Random Forest and Voting  
Classifiers: Insights from Resting State EEG Streams (2018).