



# Prediction of Diabetic Retinopathy using Naive Bayes Classifier

Divya YA

Assistant Professor, Department of Information Science and Engineering, GSSS Institute of Engineering and Technology for Women, Mysore, India

**Abstract:** A major challenge facing healthcare organizations is the provision of quality services at affordable costs. Quality service implies diagnosing patients correctly and administering treatments that are effective. Poor clinical decisions can lead to disastrous consequences which are therefore unacceptable. Hospitals must also minimize the cost of clinical tests. They can achieve these results by employing appropriate computer-based information. Most hospitals today employ some sort of hospital information systems to manage their healthcare or patient data. Clinical decisions are often made based on doctors' intuition and experience rather than on the knowledge rich data hidden in the database. The main objective of this research is to develop a Intelligent Diabetic Retinopathy Prediction System using Naive Bayes.

**Keywords:** Diabetic Retinopathy; Naive bayes; Data Mining; Classification.

## I. INTRODUCTION

Data Mining represents a process developed to examine large amounts of data routinely collected. The term also refers to a Collection of tools used to perform the process. One of the useful applications in the field of medicine is the incurable chronic disease diabetes. Data mining in healthcare is an emerging field of high importance for providing prognosis and a deeper understanding of medical data. Healthcare data mining attempts to solve real world health problems in diagnosis and treatment of diseases. Data mining algorithm is used for testing the accuracy in predicting diabetic status. Classification is one of the most frequently studied problems by DM and machine learning (ML) researchers. Classification consists of predicting certain outcome based on a given input. In order to predict the outcome, the algorithm processes a training set containing a set of attributes and the respective outcome, usually called goal or prediction attribute. The algorithm tries to discover relationships between the attributes that would make it possible to predict the outcome. Diabetes mellitus (DM) or diabetes is a group of metabolic diseases in which a person has high blood sugar. This high blood sugar produces the symptoms of frequent urination, increased thirst, and increased hunger. Untreated, diabetes can cause many complications [4].

There are three main types of diabetes mellitus:

- Type 1 DM results from the body's failure to produce insulin. This form was referred to as "insulin-dependent diabetes mellitus" (IDDM)
- Type 2 DM results from insulin resistance, a condition in which cells fail to use insulin properly, sometimes also with an absolute insulin deficiency.

This form was referred to as non insulin-dependent diabetes mellitus (NIDDM)

- Gestational diabetes is the third main form and occurs when pregnant women without a previous diagnosis of diabetes develop a high blood glucose level.

The symptoms of DM include polyurea, polydipsia, and significant weight loss among others. Diagnosis depends on blood glucose levels (fasting plasma glucose = 7.0 mmol/L). DM progression is strongly linked to several complications, mainly due to chronic hyperglycemia. It is well-known that DM covers a wide range of heterogeneous pathophysiological conditions. The most common complications are divided into micro- and macro-vascular disorders, including diabetic nephropathy, retinopathy, neuropathy, diabetic coma and cardiovascular disease. Due to high DM mortality and morbidity as well as related disorders, prevention and treatment attracts broad and significant interest. Insulin administration is the main treatment for Type 1D, although insulin is also provided in certain cases of Type 2D patients, when hyperglycemia cannot be controlled through diet, weight loss, exercise and oral medication. Current medication targets primarily a) saving one's life and alleviating the disease symptoms, and b) prevention of long term diabetic complications and/or elimination of several risk factors, thereby increasing longevity. The most common anti-diabetic agents include sulfonylurea, metformin, alpha-glucosidase inhibitor, peptide analogs, non-sulfonylurea secretagogues, etc. [6]. The majority of the present anti-diabetic agents, however, exhibit numerous side-effects. In



addition, insulin therapy is related to weight gain and hypoglycemic events. Hence, anti-diabetic drug design and discovery is of great concern and concurrently a research challenge.

The evolution of diabetes immediately increases a patient's tendency for developing a broad spectrum of irreversible complications. Complications of diabetes can be largely divided into macro and micro vascular. The macro vascular complications include peripheral arterial disease, stroke and coronary artery disease. The micro vascular complications include diabetic neuropathy, diabetic retinopathy (DR) and diabetic nephropathy. The commonness of these complications is strongly related to the type of diabetes and duration of diabetes; hence, the increasing global population, predicted climb in the proportion of adults suffering from diabetes and the changing age demographics will undoubtedly be followed by an increase in the incidence of diabetic complications [5].

**II. RELATED WORK**

Tremendous works in literature related with diabetes disease diagnosis using data mining techniques have motivated our work. The researchers in the medical field diagnose and predict the diseases in addition to providing effective care for patients [3] by employing the data mining techniques. Bayesian Network classifier was proposed to predict the persons whether diabetic or not. The dataset used is collected from a hospital, which collects the information of persons with and without diabetes [2]. Researchers in medical sector implemented the system for prediction of diabetes and treatment for the disease using different data mining techniques. This can even help doctors to target out the disease [1]

**III. METHODOLOGY**

**Naive Bayes:**

Naive Bayes classifier assumes that the presence or absence of a particular feature is unrelated to the presence or absence of any other feature, given the class variable. It can be trained very efficiently in a supervised learning. In many practical applications, parameter estimation for naive Bayes models uses the method of maximum likelihood. An advantage of naive Bayes is that it only requires a small amount of training data to estimate the parameters (means and variances of the variables)

**Naive Bayes Algorithm:**

The steps are:

Step 1: Scan the dataset

Step 2: Calculate the probability of each attribute value. [n, n\_c, m, p]

Step 3: Apply the formulae

$$P(\text{attributevalue}(a_i) / \text{subjectvalue}(v_j)) = (n\_c + mp) / n + m$$

Where:

- n = the number of training examples for which v = v<sub>j</sub>
- n\_c = number of examples for which v = v<sub>j</sub> and a = a<sub>i</sub>
- p = a priori estimate for P(a<sub>i</sub>|v<sub>j</sub>)
- m = the equivalent sample size

Step 4: Multiply the probabilities by p

Step 5: Compare the values and classify the attribute values to one of the predefined set of class.

**Table 1: Diabetic Retinopathy Dataset**

Name	Description
Duration of diabetes	1. < 15 years 2. >= 15 years
HbA1c Level	1.<7.5% 2. >=7.5%
Fasting Sugar Level	1. <=126mg/dl 2. > 126
Postprandial Sugar Level	1.<=200 mg/dl 2. >200 mg/dl
Neuropathy	1.Yes 2. No
Nephropathy	1.Yes 2. No



#### IV. RESULTS AND DISCUSSIONS

The results obtained from the given dataset classified into two classes i.e patients with diabetes and without diabetes using various data mining techniques. Based on the results demonstrated, Naïve bayes classifier provides highest accuracy 98% to predict the diseases.

Below is a summary of results. We have performed experiments using Naive Bayes algorithm. A confusion matrix is a table that is used to describe the performance of a classification model (or classifier) on a set of test data for which the true values are known. For each result a confusion matrix is presented which shows the actual samples in a particular class and the predicted class.

#### Results for Naive Bayes Algorithm:

		Predicted Class	
		Presence	Absence
Actual Class	Presence	93	01
	Absence	01	05

#### CONCLUSION

In this paper, we present prediction of diabetic retinopathy using naive bayes classifier. An early detection and treatment can reduce the risk of blindness by 95 percent. Because diabetic retinopathy often lacks early symptoms, people with diabetes should get a comprehensive dilated eye examination at least once a year. People with diabetic retinopathy need eye examinations more frequently. Hence, the objective of the project is to classify the patient who is having diabetic retinopathy disease or not. In this paper, we examine the Naive bayes classifier for classification. The results by the classifier showed good performance. Thus, the work has established a successful detection method of DR which helps to diagnose the disease.

#### REFERENCES

- [1] VelidePhani Kumar and Lakshmi Velide: A Data Mining Approach for Prediction and Treatment of Diabetes Disease, International Journal of Science Inventions Today, Volume. 3, Issue 1, January 2014.
- [2] Mukesh kumari, Dr. Rajan Vohra and Anshul arora: Prediction of Diabetes Using Bayesian Network, International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014
- [3] S.Indhumathi, Mr.G.Vijaybaskar: web based health care detection using naive bayes algorithm, International Journal of Advanced Research in Computer Engineering & Technology (IJARCET) , Volume 4 Issue 9, September 2015
- [4] Mukesh kumari, Dr. Rajan Vohra , Anshul arora: Prediction of Diabetes Using Bayesian Network, International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014
- [5] Huda Mirza Saifuddin and H.C.Vijayalakshmi: Prediction of diabetic retinopathy using multi layer perceptron, International Journal of Advanced Research (2016), Volume 4, Issue 6.
- [6] Ioannis Kavakiotis, Olga Tsave, Athanasios Salifoglou, Nicos Maglaveras, Ioannis Vlahavas, Ioanna Chouvarda: Machine learning and data mining methods in diabetes research, Computational and Structural Biotechnology Journal, Volume 15, 2017
- [7] Dr. M. Renuka Devi, J. Maria Shyla: Analysis of Various Data Mining Techniques to Predict Diabetes Mellitus, International Journal of Applied Engineering Research ISSN 0973-4562 Volume 11, Number 1 (2016)