# Title: Diabetic Medication and Patient Re-admission prediction

|  |  |  |
| --- | --- | --- |
| First name | Last Name | IIT Email |
| Prashant | Raina | [praina@hawk.iit.edu](mailto:praina@hawk.iit.edu) |
| Vaidehi | Rathkanthiwar | vrathkanthiwar@hawk.iit.edu |
| Utkarsha | Vidhale | [uvidhale@hawk.iit.edu](mailto:uvidhale@hawk.iit.edu) |

1. **Introduction**

The management of sugar level in the hospitalized patient has a significant bearing on outcome, in terms of both morbidity and mortality. This recognition has led to the development of formalized protocols in the intensive care unit (ICU) setting with rigorous glucose targets in many institutions. However, the same cannot be said for most non-ICU inpatient admissions. This analysis of a large clinical database is to be undertaken to examine historical patterns of diabetes care in patients admitted to a US hospital and to inform future directions which might lead to improvements in patient’s medical condition and help to save medical resources and valuable time of medical staff.

Our application will provide general idea of patient’s medical necessity based on diabetic prescription and number of times patient is re-admitted in hospital. Since, there has always been shortage of resources in medical industry whether it is hospital bed, medicines and other equipment due to large number of patients, so it is always advisable to keep the tab of patients that might use the resources in future. Our model will provide the predictions in terms of whether patient will be prescribed diabetic medication or not can really help the doctors and medical staff to keep the proper tab of patient’s health and avoid diabetic condition. This model will also determine if patient needs to be readmitted in hospital based on clinical history, medication and other factors which can help the medical staff to properly maintain and utilize medical resources as per the necessity and availability.

1. **Data Sets**

The dataset represents 10 years (1999-2008) of clinical care at 130 US hospitals and integrated delivery networks. It is sourced from UCI Machine Learning Repository and has been prepared to analyze factors related to readmission as well as other outcomes pertaining to patients with diabetes. (<http://archive.ics.uci.edu/ml/datasets/Diabetes+130-US+hospitals+for+years+1999-2008>).

The data contains more than 1,00,000 instances and 50 attributes. The dataset is multivariate in terms of its characteristics, whereas the attributes are numerical and nominal.

Attribute list and their datatype (Numerical or Nominal)

encounter\_id : numerical

patient\_nbr : numerical

race : nominal

gender : nominal

age : nominal

weight : nominal

admission\_type\_id : numerical

discharge\_disposition\_id : numerical

admission\_source\_id : numerical

time\_in\_hospital : numerical

payer\_code : nominal

medical\_speciality : nominal

num\_lab\_procedures : numerical

num\_medications : numerical

number\_outpatient : numerical

number\_emergency : numerical

numbert\_inpatient : numerical

diag\_1 : nominal

diag\_2 : nominal

diag\_3 : nominal

number\_diagnoses : numerical

max\_glu\_serum : nominal

A1Cresult : nominal

metformin : nominal

repaglinide : nominal

nateglinide : nominal

chlorpropamide : nominal

glimepiride : nominal

acetohexamide : nominal

glipizide : nominal

glyburide : nominal

tolbutamide : nominal

pioglitazone : nominal

rosiglitazone : nominal

acarbose : nominal

miglitol : nominal

troglitazone : nominal

tolazamide : nominal

examide : nominal

citoglipton : nominal

insulin : nominal

glyburide-metformin : nominal

glipizide-metformin : nominal

glimepiride-metformin : nominal

metformin-rosiglitazone : nominal

metformin-pioglitazone : nominal

change : nominal

diabetesMed : nominal

readmitted : nominal

1. **Research Problems**

Based on the above dataset, following are some of the interesting research problems:

1. Based on the clinical condition, patient’s physical features and medical history of patient, we are going to predict whether the patient should be treated with diabetic’s medication or not. This can be helpful for both patients and doctors as they can keep the tab of the health of their patients with this prediction model and take some early steps to avoid the condition of diabetic medication.
2. In order to improve patient’s safety and doctor’s valuable time, we are going to determine whether the patient will be required to readmit in the hospital within 30 days, after 30 days or never, based on the clinical history, diabetics prescription and other factors. This can provide the proper timeline of patient’s medical history and will help the clinics and hospitals to utilize proper medical resources based on availability of patient and the critical condition of patient, thus avoiding the chaos.
3. **Potential Solutions**

Following are some of the feasible solutions that we may use in our data model:

1. We will use binary classification technique to determine whether the patient is given diabetic medication or not, since this label is classified into Yes or No. We will use Random Forest classifier for the above classification technique as it is the best algorithm for working with dataset having large number of attributes and it handles unbalanced data in a better way.

1. The number of readmission times are categorized into three classes as no re-admission, less than 30 days and more than 30 days. Thus, we will use various multi classification techniques such as K means clustering and Random Forest classifier to determine whether the patient is re-admitted or not as they are best algorithms for working with large number of attributes and can handle unbalanced data in a better way.
2. **Evaluations**

We will take following steps in evaluating our model:

1. The proposed dataset has some noisy and missing data which will be treated with proper preprocessing techniques like missing value imputation, normalization methods or standardization methods.
2. As there are large number of instances (approximately 0.1 million), hold out evaluation will be used to obtain the training and test dataset. We will split the data into 80-20 ratio i.e training set (80%) and test set (20%) and will start building the model using training data.
3. We will build different models and evaluate them using accuracy score or confusion matrix to get the best working algorithm.
4. **Expected Outcomes**

Following are some of the outcomes that we will get after building our models:

1. We will be able to determine whether the patient should be treated with diabetic’s medication or not based on the clinical condition, patient’s physical features and medical history of patient.

This would help us in knowing if the patient is on the verge of cure or not.

1. We will be able to determine whether the patient will be required to readmit in the hospital within 30 days, after 30 days or never based on the clinical history, diabetics prescription and other factors. This would help us in knowing how critical the diabetic condition for the patient is as well as it will help the clinics and hospitals to utilize proper medical resources based on availability of patients and their critical condition.