

# **Using Machine Learning for Prediction of Early Readmission of Diabetic Patients**

## **Project Abstract**

**CIND 820- Big Data Analytics Project**

**Supervised by: Ceni Babaoglu**

**Presented by: Nehal Gamal Mohamed (501278190)**

**Date of Submission: 5/21/2024**

Diabetes management poses a significant challenge in healthcare, with patient readmission within 30 days of discharge serving as a critical metric for assessing care quality. Despite advances in preventive interventions, many diabetic patients experience readmissions due to suboptimal glycemic control and inadequate care.

This study will try to predict the likelihood of early readmission (within 30 days) for patients diagnosed with diabetes using clinical data collected over a ten-year period (1999-2008) from 130 US hospitals and integrated delivery networks.

The study objective is to answer the following questions:

- By using machine learning models can we accurately predict early readmission of diabetic patients?
- What are the patient and hospital factors which strongly influence early readmission?
- By using predictive models how can we improve diabetes management and reduce readmission rates?

The dataset has patient records from diabetic encounters, including demographic information, medical history, admission details, laboratory results, medications given, and hospital outcomes. It has 101,766 instances and 47 features, the data is multivariate, consisting of categorical and integer variables.

The study will commence with preparing the dataset, then employing classification techniques to predict early readmission outcomes based on patient and hospital features. Machine learning algorithms including logistic regression, decision trees, and random forests will be utilized.

After splitting the dataset and training the model, the performance will be evaluated using standard metrics such as accuracy, precision, recall and F1 score. Moreover, feature selection and dimensionality reduction techniques will be applied to identify the most informative factors and enhance model performance. Additionally, clustering algorithms will be explored to check if there are hidden patterns that could reveal any patient subgroups with different readmission risks.

Python programming will be used for implementation, to take advantage of various python libraries and tools.

## References & Citation:

### Dataset Source

<https://archive.ics.uci.edu/dataset/296/diabetes+130-us+hospitals+for+years+1999-2008>

### Introductory Paper

Strack, B., DeShazo, J. P., Gennings, C., Olmo, J. L., Ventura, S., Cios, K. J., & Clore, J. N. (2014). Impact of HbA1c measurement on hospital readmission rates: Analysis of 70,000 clinical database patient records. *BioMed Research International*, 2014, 781670–781611.  
<https://doi.org/10.1155/2014/781670>

URL: <https://www.hindawi.com/journals/bmri/2014/781670/>