**Hyperparameter Database**

*Michelle Pradeep |* [*pradeep.m@husky.neu.edu*](mailto:pradeep.m@husky.neu.edu)

*Seemanthini Jois |* [*jois.s@husky.neu.edu*](mailto:jois.s@husky.neu.edu)

*Suraksha Jadhav**|* [*jadhav.su@husky.neu.edu*](mailto:jadhav.su@husky.neu.edu)

***Abstract*--*Management of hyperglycemia in hospitalized patients has a significant bearing on outcome, in terms of both morbidity and mortality. However, there are few national assessments of diabetes care during hospitalization which could serve as a baseline for change. This analysis of a large clinical database(74 million unique encounters corresponding to 17 million unique patients was under taken to provide such an assessment and to find future directions which might lead to improvements in patient safety. Almost 70,000 in patient diabetes encounters were identified with sufficient detail for analysis. Multi variable logistic regression was used to fit the relationship between the measurement of HbA1c and early readmission while controlling or covariates such as demographics, severity and type of the disease, and type of admission. Results show that the measurement of HbA1c was performed in frequently (18.4%) in the inpatient setting. The statistical model suggests that the relationship between the probability of readmission and the HbA1c measurement depends on the primary diagnosis. The data suggest further that the greater attention to diabetes reflected in HbA1c determination may improve patient outcomes and lower cost of inpatient care.***

**Keywords**

Diabetes, Hyperparameters, Distributed Random Forest (DRF), Generalized Linear Model (GLM), Gradient Boosting Machine (GBM). Naïve Bayes Classifier, Stacked Ensembles, Xgboost and Deep Learning Models (Neural Networks)

1. **Introduction**

It is increasingly recognized that the management of hyperglycemia 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. Rather, anecdotal evidence suggests that inpatient management is arbitrary and often leads to either no treatment at all or wide fluctuations in glucose when traditional management strategies are employed. Although data are few, recent controlled trials have demonstrated that protocol driven in patient strategies can be both effective and safe. As such , implementation of protocols in the hospital setting is now recommended.

However, there are few national assessments of diabetes care in the hospitalized patient which could serve as a baseline for change. The present analysis of a large clinical data base was under taken to examine historical patterns of diabetes care in patients with diabetes admitted to a US hospital and to inform future directions which might lead to improvements in patient safety. In particular, we examined the use of HbA1c as a marker of attention to diabetes care in a large number of individuals identified as having a diagnosis of diabetes mellitus.

We hypothesize that measurement of HbA1c is associated with a reduction in readmission rates in individuals admitted to the hospital. Databases of clinical data contain valuable but heterogeneous and difficult data in terms of missing values, incomplete or inconsistent records , and high dimensionality understood not only by number of features but also their complexity.

Additionally, analyzing external data is more challenging than analysis of results of a carefully designed experiment or trial, because one has no impact on how and what type of information was collected. Nonetheless, it is important to utilize these huge amounts of data to find new information/knowledge that is possibly not available anywhere.

1. **Dataset**

We picked a publicly available dataset from UCI repository containing de-identified diabetes patient encounter data for 130 US hospitals (1999–2008) containing 101,766 observations over 10 years. The dataset has over 50 features including patient characteristics, conditions, tests and 23 medications. Only diabetic encounters are included (i.e. at least one of three primary diagnosis was diabetes).

1. **Method**

Hyperparameters are parameters that are specified prior to running machine learning algorithms that have a large effect on the predictive power of statistical models. Knowledge of the relative importance of a hyperparameter to an algorithm and its range of values is crucial to hyperparameter tuning and creating effective models. The hyperparameter database is a public resource with algorithms, tools, and data that allows users to visualize and understand how to choose hyperparameters that maximize the predictive power of their models. The hyperparameter database is created by running millions of hyperparameter values, over thousands of public datasets and calculating the individual conditional expectation of every hyperparameter on the quality of a model. Currently, the hyperparameter database analyzes the effect of hyperparameters on the following algorithms: Distributed Random Forest (DRF), Generalized Linear Model (GLM), Gradient Boosting Machine (GBM). Naïve Bayes Classifier, Stacked Ensembles, Xgboost and Deep Learning Models (Neural Networks). The hyperparameter database also uses these data to build models that can predict hyperparameters without search and for visualizing and teaching statistical concepts such as power and bias/variance tradeoff.

1. **Code and Documentation**

**GitHub Link:** The complete code with documentation can be found on the below link :

1. **Results**
2. **Discussion**
3. **Conclusions**
4. **Acknowledgment**

We are thankful to our Prof. Nik Bear Brown, Assistant Professor, Northeastern University, Boston, MA for his valuable guidance, encouragement and co-operation during the implementation of this project.

1. **References**

[1] Impact of HbA1c Measurement on Hospital Readmission Rates: Analysis of 70,000 Clinical Database Patient Records

[2]<https://github.com/skunkworksneu/Projects/blob/master/Hyperparameter%20Database.pdf>