Prediction of Gestational diabetes using machine learning

Name: Shoaib Anwar Email: sanwar1@crimson.ua.edu

*Abstract*—

*1-2 sentences: How significant of the research questions you are proposing and any existing challenges related to your work? Pretty much like summarizing sections I-III*

*3-4 sentences: Any novelty in your work, or outline how you implement your work. Like, summarize section IV.*

*5-6 sentences: Summarize the most promising results in Section V and explain the implications or potential applications of the results discussed in Section VI.*

Keywords— gestational diabetes, machine learning, early prediction, AutoML, logistic regression

# Introduction (*Heading 1*)

Gestational diabetes is a type of diabetes that can develop during pregnancy in women who don't already have diabetes. Every year, 2% to 10% of pregnancies in the United States are affected by gestational diabetes. Managing gestational diabetes will help make sure you have a healthy pregnancy and a healthy baby. Early prediction of gestational diabetes may prepare to-be mothers to have proper care and diet to deal with the issue.

# Research Questions

All forms of diabetes depend on both physical and genetic attributes. Therefore, it would be very easy to predict whether a woman can develop diabetes during pregnancy. The research questions on which this study is based, are -

“Can gestational diabetes or diabetes during pregnancy be predicted early using physical attributes by machine learning? What sign indicates the patient would have diabetes?”

# Related Work

Several studies have been made on predicting gestational diabetes. Zheng et.al. develop an ML model to the predict risk of gestational diabetes mellitus from 8 to 20 weeks of gestation in Chinese women [1]. Instead of using several physical attributes, the model can be developed focusing on specific physical attributes [2]. They chose a dataset of a total of 4771 pregnant women with early gestation diabetes. This study is based on the physical attribute. Models can be developed based on genetic attributes too [3]. All of these models used logistic regression to analyze.

# Methods

The aim of this study is to analyze the hypothesis mentioned in section II. To do so, the first step is to acquire suitable and reliable data, then, clean the unnecessary and false information. After that, the next step is to build a model with appropriate hyperparameters and train the model with the data and finally, deploy the model for reuse. The following subsections will describe these steps in detail.

## Data Acquisition

I searched for the data which should contain the physical attributes related to pregnant women. The dataset must have information on the diabetes of the selected patient. I searched several websites from where I could download the datasets which were publicly available. I collected the data from the Kaggle website. This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. It contains data on 768 pregnant females whether they had diabetes or not and the physical factors that could influence having diabetes.

Source:

* Collected from: <https://www.kaggle.com/datasets/mathchi/diabetes-data-set>
* Original source: National Institute of Diabetes and Digestive and Kidney Diseases
* Donor of the dataset: Vincent Sigillito ([vgs@-aplcen.apl.jhu.edu](mailto:vgs@-aplcen.apl.jhu.edu)), Research Center, RMI Group Leader Applied Physics Laboratory, The Johns Hopkins University Johns Hopkins Road

Format: CSV

## Data Cleaning and Labeling

Features: It has a total of 9 features.

• Pregnancies: Number of times pregnant

• Glucose: Plasma glucose concentration 2 hours in an oral glucose tolerance test

• BloodPressure: Diastolic blood pressure (mm Hg)

• SkinThickness: Triceps skin fold thickness (mm)

• Insulin: 2-Hour serum insulin (mu U/ml)

• BMI: Body mass index (weight in kg/(height in m)^2)

• DiabetesPedigreeFunction: Diabetes pedigree function

• Age: Age (years)

• Outcome: Class variable (0 or 1).

## Feature Engineering

Please describe how you implement feature extraction and representation.

## Model Development

Please describe how you develop the models, including SVM, Decision Trees, Linear Regression, and Neural Networks.

## Model Training

Please describe how you train the model, especially any optimization strategies you adopt for hyperparameters and model parameters.

## Model Evaluation and Validation

Please describe the strategies you used for evaluation and validation of the trained models.

## Model Deployment

Please describe how you deploy the models to web.

# Results

Please show your results as below.

## Mode Performance in Evaluation

#### Please show the prediction performance of the models during evaluation and validation.

## Feature Importance

Please show how the features contribute to the model performance.

## Runtime Monitoring

Find other 2 classmates and let them use your web-based models. Each of them should create a testing data sample to test your web-based model performance. Please describe your web-based model’s testing performance here.

# Discussion

## Answers to Research Questions

Did your results answer the research questions. If yes, how much is being answered? If no, why?

## Limitations of your research.

Please explain any limitations of your research

## Lessons Learned

Please describe what kind of lessons you learned from the final project.

##### Acknowledgment

Please acknowledge your classmates and any others helped you during the research process..

##### References

The template will number citations consecutively within brackets [1]. The sentence punctuation follows the bracket [2]. Refer simply to the reference number, as in [3]—do not use “Ref. [3]” or “reference [3]” except at the beginning of a sentence: “Reference [3] was the first ...”

Number footnotes separately in superscripts. Place the actual footnote at the bottom of the column in which it was cited. Do not put footnotes in the abstract or reference list. Use letters for table footnotes.

Unless there are six authors or more give all authors’ names; do not use “et al.”. Papers that have not been published, even if they have been submitted for publication, should be cited as “unpublished” [4]. Papers that have been accepted for publication should be cited as “in press” [5]. Capitalize only the first word in a paper title, except for proper nouns and element symbols.

For papers published in translation journals, please give the English citation first, followed by the original foreign-language citation [6].

1. Zheng, T., Ye, W., Wang, X., Li, X., Zhang, J., Little, J., Zhou, L., Zhang, L., 2019. A simple model to predict risk of gestational diabetes mellitus from 8 to 20 weeks of gestation in Chinese women. BMC Pregnancy and Childbirth 19.. doi:10.1186/s12884-019-2374-8.
2. Ramezani S, Doulabi MA, Saqhafi H, Alipoor M. Prediction of Gestational Diabetes by Measuring the Levels of Pregnancy Associated Plasma Protein-A (PAPP-A) During Gestation Weeks 11-14. J Reprod Infertil. 2020 Apr-Jun;21(2):130-137. PMID: 32500016; PMCID: PMC7253940.J. Clerk Maxwell, A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.
3. Wu, Q., Chen, Y., Zhou, M., Liu, M., Zhang, L., Liang, Z., Chen, D., 2022. An early prediction model for gestational diabetes mellitus based on genetic variants and clinical characteristics in China. Diabetology & Metabolic Syndrome 14.. doi:10.1186/s13098-022-00788-y
4. R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.
5. Y. Yorozu, M. Hirano, K. Oka, and Y. Tagawa, “Electron spectroscopy studies on magneto-optical media and plastic substrate interface,” IEEE Transl. J. Magn. Japan, vol. 2, pp. 740–741, August 1987 [Digests 9th Annual Conf. Magnetics Japan, p. 301, 1982].
6. M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

**IEEE conference templates contain guidance text for composing and formatting conference papers. Please ensure that all template text is removed from your conference paper prior to submission to the conference. Failure to remove template text from your paper may result in your paper not being published.**