


# Diabetes Risk: The Prediction of Diabetes Risk Through Deep Learning

By: Katherine Nguyen



# Agenda

- Outline/Agenda
- Introduction
- Approach
  - Data
  - Data Cleaning
  - Models
  - Model Analysis
- Results
- Conclusion



# Introduction

- **Problem**

- How can we predict diabetes based on potential risk factors of diabetes?

- **Purpose**

- To understand if diabetes risk factors that can help predict whether an individual will acquire diabetes based on those risk factors in the future

- **Why Is it Important?**

- To potentially acknowledge risk factors that may contribute to predicting diabetes
- Raise awareness



## Approach: Data

- **Dataset:** "Diabetes Risk Prediction"
  - Diabetes Database: <https://www.kaggle.com/datasets/rcratos/diabetes-risk-prediction>
  - Categorical Data; Binary

## Approach: Data

|   | Age | Gender | Polyuria | Polydipsia | sudden weight loss | weakness | Polyphagia | Genital thrush | visual blurring | Itching | Irritability | delayed healing |
|---|-----|--------|----------|------------|--------------------|----------|------------|----------------|-----------------|---------|--------------|-----------------|
| 0 | 40  | Male   | No       | Yes        | No                 | Yes      | No         | No             | No              | Yes     | No           | Yes             |
| 1 | 58  | Male   | No       | No         | No                 | Yes      | No         | No             | Yes             | No      | No           | No              |
| 2 | 41  | Male   | Yes      | No         | No                 | Yes      | Yes        | No             | No              | Yes     | No           | Yes             |
| 3 | 45  | Male   | No       | No         | Yes                | Yes      | Yes        | Yes            | No              | Yes     | No           | Yes             |
| 4 | 60  | Male   | Yes      | Yes        | Yes                | Yes      | Yes        | No             | Yes             | Yes     | Yes          | Yes             |



## Approach: Data Cleaning

- **Data Cleaning :**
  - Simplified the Data
    - Convert labels into binary values (e.g. Yes = 1, No = 0)
    - Removed unnecessary features for prediction
    - Remove unimportant features

## Approach: Data

|                       | Age | Gender | Polyuria | Polydipsia | sudden weight loss | weakness | Polyphagia | Genital thrush | visual blurring | Itching | Irritability | delayed healing | partial paresis | muscle stiffness | Alopecia | Obesity | class |
|-----------------------|-----|--------|----------|------------|--------------------|----------|------------|----------------|-----------------|---------|--------------|-----------------|-----------------|------------------|----------|---------|-------|
| 0                     | 40  | 1      | 0        | 1          | 0                  | 1        | 0          | 0              | 0               | 1       | 0            | 1               | 0               | 1                | 1        | 1       | 1     |
| 1                     | 58  | 1      | 0        | 0          | 0                  | 1        | 0          | 0              | 1               | 0       | 0            | 0               | 1               | 0                | 1        | 0       | 1     |
| 2                     | 41  | 1      | 1        | 0          | 0                  | 1        | 1          | 0              | 0               | 1       | 0            | 1               | 0               | 1                | 1        | 0       | 1     |
| 3                     | 45  | 1      | 0        | 0          | 1                  | 1        | 1          | 1              | 0               | 1       | 0            | 1               | 0               | 0                | 0        | 0       | 1     |
| 4                     | 60  | 1      | 1        | 1          | 1                  | 1        | 1          | 0              | 1               | 1       | 1            | 1               | 1               | 1                | 1        | 1       | 1     |
| ...                   | ... | ...    | ...      | ...        | ...                | ...      | ...        | ...            | ...             | ...     | ...          | ...             | ...             | ...              | ...      | ...     | ...   |
| 515                   | 39  | 0      | 1        | 1          | 1                  | 0        | 1          | 0              | 0               | 1       | 0            | 1               | 1               | 0                | 0        | 0       | 1     |
| 516                   | 48  | 0      | 1        | 1          | 1                  | 1        | 1          | 0              | 0               | 1       | 1            | 1               | 1               | 0                | 0        | 0       | 1     |
| 517                   | 58  | 0      | 1        | 1          | 1                  | 1        | 1          | 0              | 1               | 0       | 0            | 0               | 1               | 1                | 0        | 1       | 1     |
| 518                   | 32  | 0      | 0        | 0          | 0                  | 1        | 0          | 0              | 1               | 1       | 0            | 1               | 0               | 0                | 1        | 0       | 0     |
| 519                   | 42  | 1      | 0        | 0          | 0                  | 0        | 0          | 0              | 0               | 0       | 0            | 0               | 0               | 0                | 0        | 0       | 0     |
| 520 rows × 17 columns |     |        |          |            |                    |          |            |                |                 |         |              |                 |                 |                  |          |         |       |

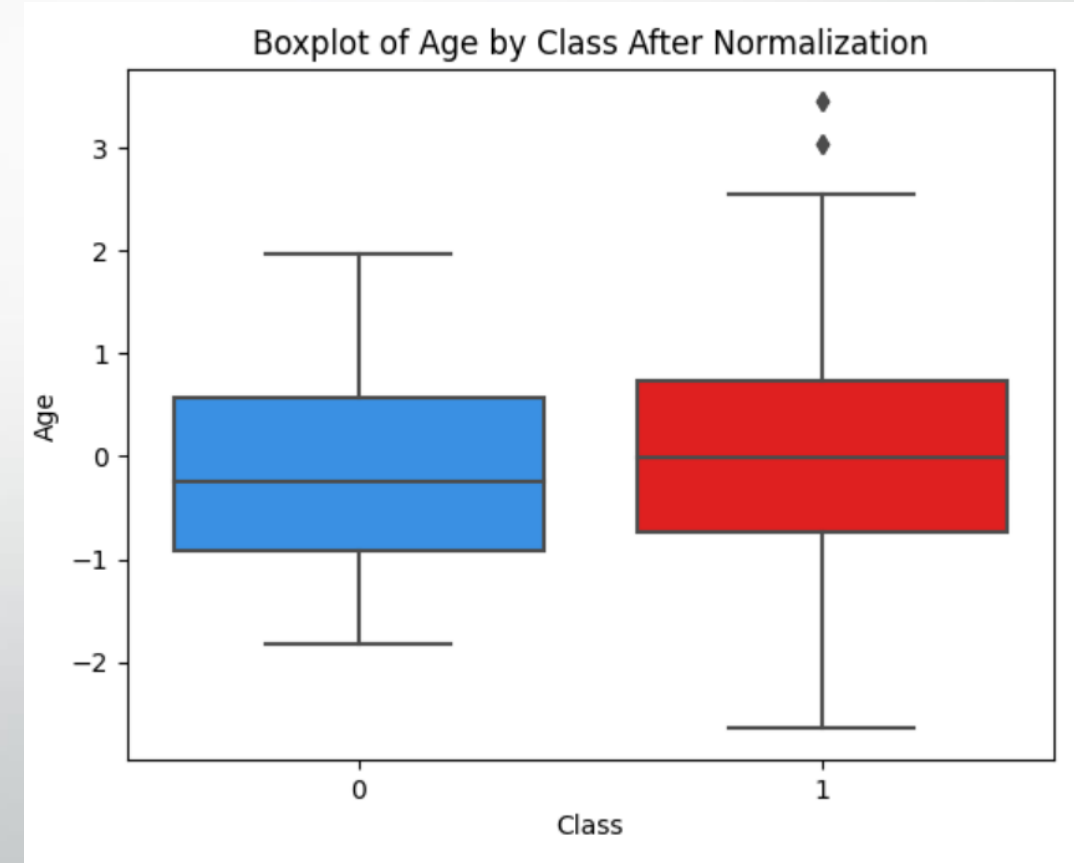
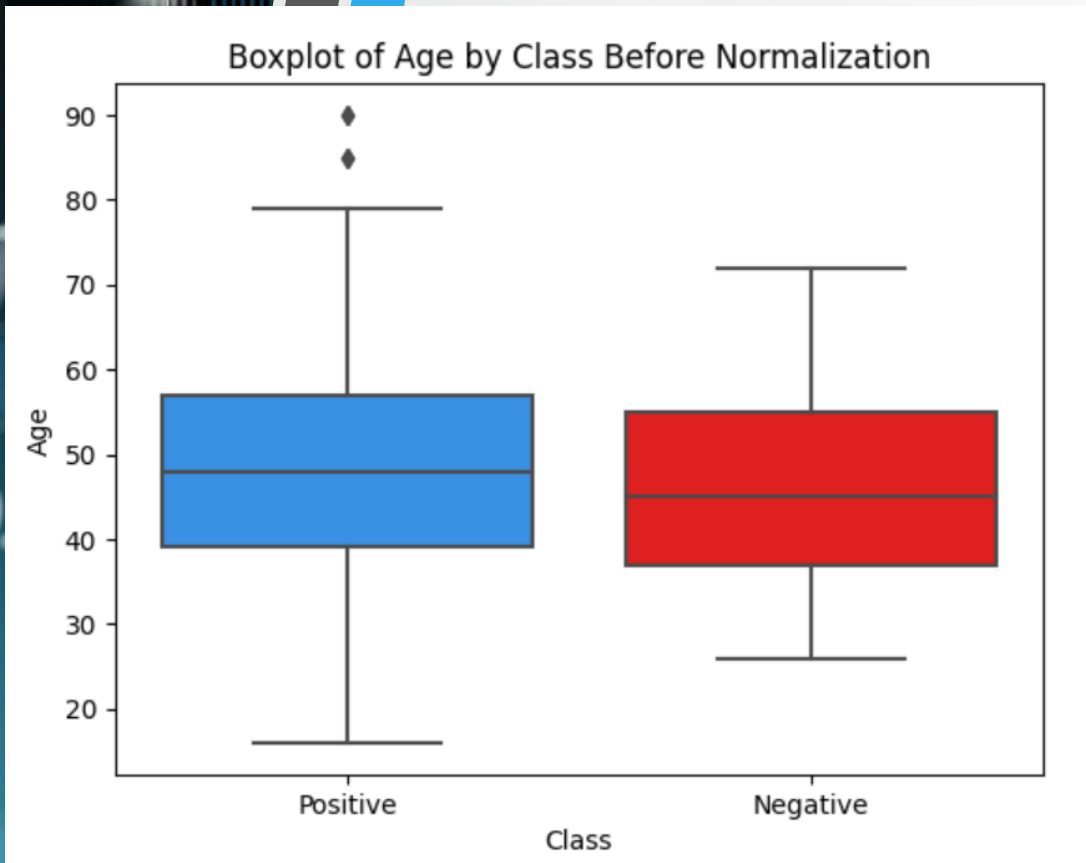


## Approach: Exploratory Data Analysis

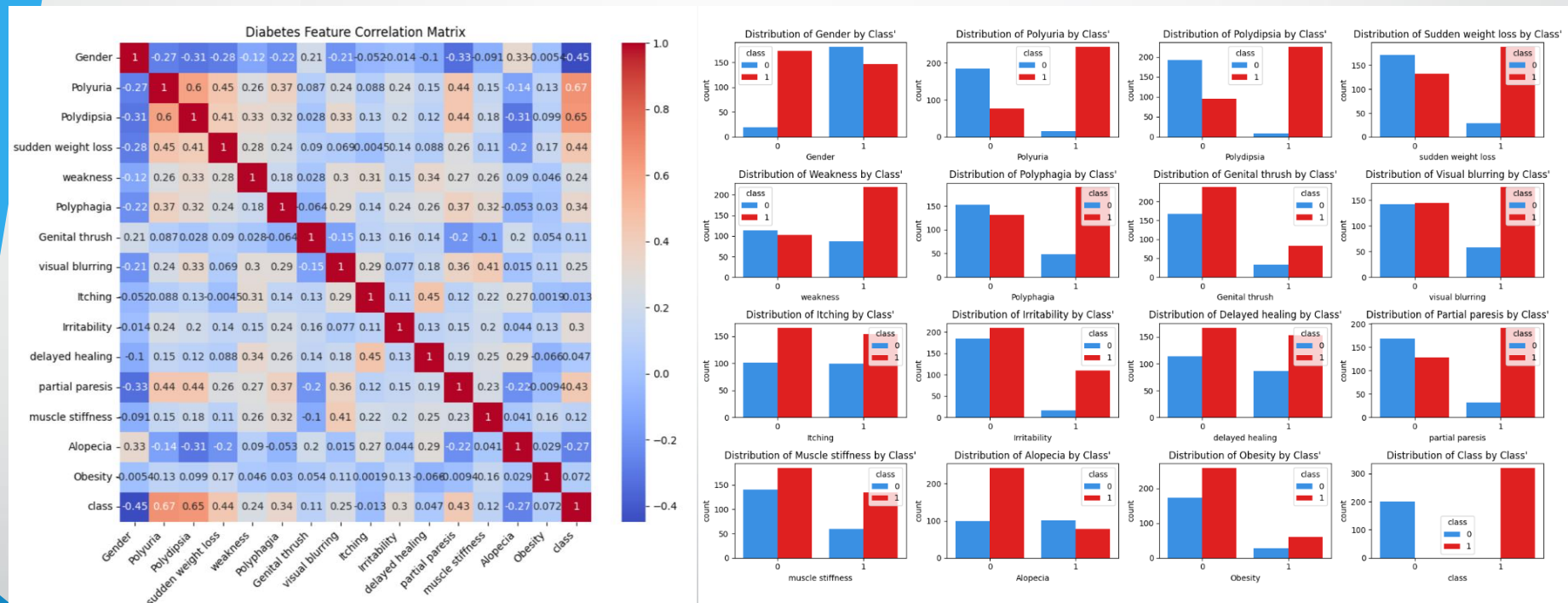
- **Exploratory Data Analysis :**
  - Used Bar Plots to show classifications of diabetes in respect to other features



# Approach: Data Cleaning



# Approach: Data Cleaning



## Approach: Data Cleaning

|     | Polyuria | Polydipsia | Gender | sudden weight loss | partial paresis |
|-----|----------|------------|--------|--------------------|-----------------|
| 0   | 0        | 1          | 1      | 0                  | 0               |
| 1   | 0        | 0          | 1      | 0                  | 1               |
| 2   | 1        | 0          | 1      | 0                  | 0               |
| 3   | 0        | 0          | 1      | 1                  | 0               |
| 4   | 1        | 1          | 1      | 1                  | 1               |
| ... | ...      | ...        | ...    | ...                | ...             |
| 515 | 1        | 1          | 0      | 1                  | 1               |
| 516 | 1        | 1          | 0      | 1                  | 1               |
| 517 | 1        | 1          | 0      | 1                  | 1               |
| 518 | 0        | 0          | 0      | 0                  | 0               |
| 519 | 0        | 0          | 1      | 0                  | 0               |

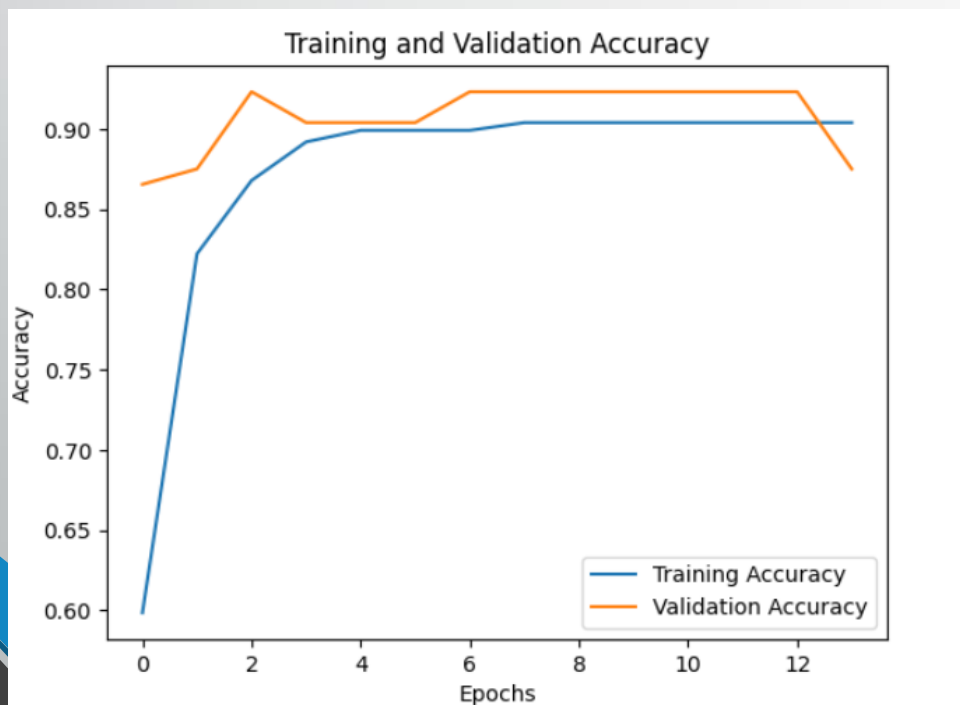


## Approach: Models

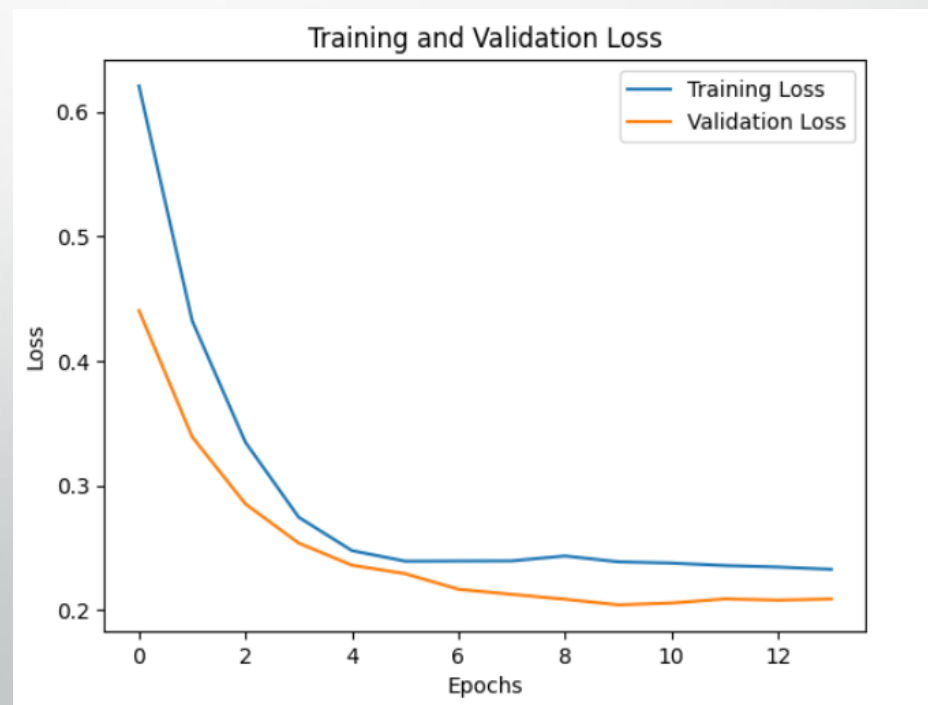
- **Deep Learning:**
  - Convolutional Neural Networks (CNN)
    - Predicts and learns based on layers in relation to sequential variables.
    - Failed to Run
  - Forward Neural Networks
    - Predicts and learns about data with independent features

# Approach: Models Analysis

## FNN Accuracy Plot

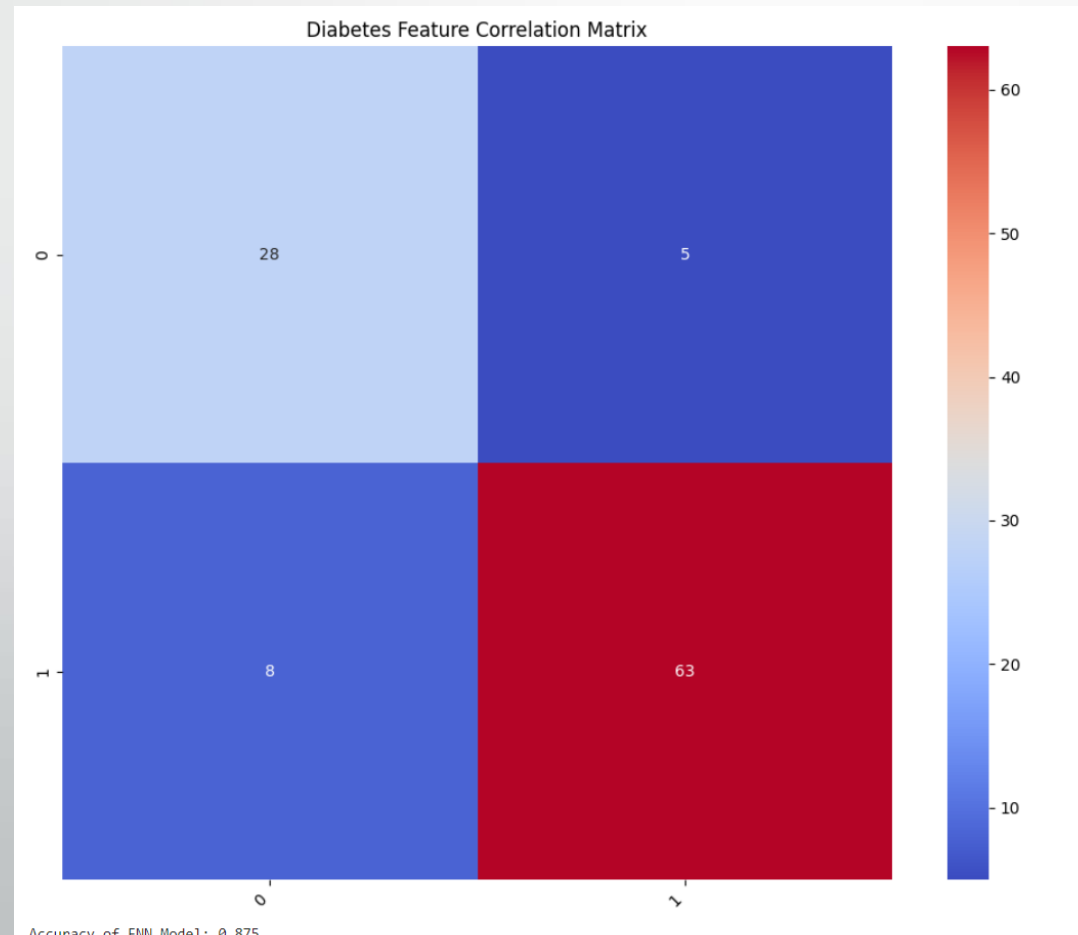


## FNN Loss Plot



# Approach: Models Analysis

## Confusion Matrix





# Results

- **Evaluation:**

- CNN
  - Does Not Apply to this Data
- FNN
  - Somewhat accurate, needs improvement
  - Diabetes are predictable based on the risk factors of
    - Polyuria
    - Polydipsia
    - Gender
    - sudden weight loss
    - partial paresis



## Conclusion

- **CNN** does not work with this kind of data but **FNN** does
- **Diabetes** is predictable via deep learning, especially for
  - Polyuria
  - Polydipsia
  - Gender
  - sudden weight loss
  - partial paresis