Software Architecture Project -002

Code Smell detection using Machine Learning and Deep Learning

In this project we are creating a code smell classifier using machine learning and deep learning. First We have taken 2 csv files as input code\_corpus.csv and dictionary.csv.These files are taken from github. These csv files contain projects on which we need to find a bad smell.

We have divided this project into the following three parts.

.

1. **Creating features.csv using Deep learning:**

In this part we have taken code\_corpus.csv and dictionary.csv files and created a deep learning model to find features which are responsible for bed smells. Here we are taking 4 type of code smell into consideration:

1. Data class
2. Feature envy
3. God class
4. Long Method

1. **Applying machine learning to classify smell:**

In this part we are applying machine learning to classify the bed smells into 4 types. For this we are using a 4 machine learning models which are as follows:’

1. Decision Tree
2. Random Forest
3. Support Vector Machine
4. Logistic regression

For each model we calculated training and test accuracy as well as confusion matrix which can be seen in the following table.

| SR.No | Classifier | Training Acc. | Testing Acc. |
| --- | --- | --- | --- |
| 1 | Decision Tree | 48.59474% | 41.304347% |
| 2 | Random Forest | 99.3653% | 15.5797% |
| 3 | Support Vector Machine | 30.6436% | 31.1594% |
| 4 | Logistic Regression | 51.4052% | 31.5217% |

1. **Applying Deep Learning to Classify Smell:**

In this part we applied 2 deep learning techniques called Gated Recurrent Unit(GRU) and (Long Short Term Memory) LSTM.LSTM ha GRU performs better then LSTM here.