Lending Club Loan Data Analysis.

Project 1

DESCRIPTION

Create a model that predicts whether or not a loan will be default using the historical data.

 Problem Statement:

For companies like Lending Club correctly predicting whether or not a loan will be a default is very important. In this project, using the historical data from 2007 to 2015, you have to build a deep learning model to predict the chance of default for future loans. As you will see later this dataset is highly imbalanced and includes a lot of features that makes this problem more challenging.

**Analysis to be done:** Perform data preprocessing and build a deep learning prediction model.

**Following actions are actually performed:**

1. **Data Exploration**

Learned that the dataset contains 9578 records, out of which 1533 are defaulted loans. The dataset has 1 response and 12 features. Out of the features, 1 is categorical.

Pairplot is used to check if there is any observable relationship between features and also the response. No clear relationship pattern is observed.

1. **Data Wrangling**

Convert categorical data into numerical values and check correlation between different features.

PCA is also used to perform dimension reduction.

1. **Data Modeling**

Three models using Keras with Tensorflow backend were constructed based on 3 stages of data process. 1st one is utilizing the original dataset, and 2nd one is utilizing the data after removing attributes with high correlation. The last one is utilizing the data after the PCA.

Similar level of accuracy is achieved by all the three models, which is illustrating the robustness of Keras to consume unprocessed data for training.