IMSE 685 Manufacturing Information Systems
Midterm Exam II

Credit Risk Model (EDA & ML Model Selections)

Exam Notebook: Final_Exam_S23.ipynb

Datasets: lending_club_info.csv Download
lending_club_info.csv, lending_club_loan_two.csvDownload lending_club_loan_two.csv

Some reference materials:

- https://data.world/jaypeedevlin/lending-club-loan-data-2007-11Links to an external site.
- https://towardsdatascience.com/mapping-inequality-in-peer-2-peer-lending-using-geopandas-part-1-b8c7f883d1baLinks to an external site.
- https://github.com/jgillick/LendingClubLinks to an external site.
- https://www.dataquest.io/blog/machine-learning-preparing-data/Links to an external site.
- https://www.kaggle.com/wendykan/lending-club-loan-dataLinks to an external site.
- https://towardsdatascience.com/predicting-loan-repayment-5df4e0023e92Links to an external site.

Required Parts:

1) **Part 1** (EDA, 50 points) Base notebook provided. Finish the EDA portion of blank cells in the attached Jupyter Notebook. What is the number of features you end up with?
2) **Part 2** (60 points) Using SVM and Random Forest algorithms (30 points each) that we have taught with/without LDA to perform the same classification. Add this part to your notebook and have a proper section title of "Part 2."
3) **Part 3** (120 points) Finished the Notebook using the Neural Network model. Add this part to your notebook and have a section title of "Part 3."
    a) (Fully Connected Network, 10 points) Set up a neural network with 54(ReLu) × 27(ReLu) × 14(ReLu) × 1(sigmoid, output), with loss function = "binary_crossentropy".
    b) (30 points) First, train the neural network you built with 30 epochs with batch_size=1024. Plot the history of the loss function values.
    c) (Early Stop, 30 points) Then, set up an early stop mechanism to stop your training when "val_loss" is not improving after 5 epochs. Plot the history of the loss function values.
    d) (Drop-Out Layers + Early Stop, 30 points) Set up a neural network with 78(ReLu) × 39(ReLu) × 19(ReLu) × 1(sigmoid, output), with loss function = "binary_crossentropy". Then, add drop-out layers of 50% after 1st, 2nd, and 3rd layers, and retrain your network with "Early Stop" mechanism. Plot the history of the loss function values.
    e) (20 points) Present both the confusion matrix and classification report for parts b), c), and d) and comment and compare on their results.
4) **Part 4** (40 points) Compare your results between Part 2 and Part 3 using various performance metrics (e.g., classification_report, confusion_matrix, etc.) and justify your recommendations. Add this part to your notebook and have a section title of "Part 4."

Notices:

- one-week take-home exam
- Work in max. of 4 students in a group
- Any suspected violations of the K-State honor pledge will be forwarded directly to University Honor's Committee, and you will receive an "XF" grade for the entire class until you've been proved innocent.
- The instance with your name and involved classes) will be published on the University Honor System website for more than 10 years.
- Your potential employers may be notified of your suspected violation of academic dishonesty.