Maaloomatiia: Data Scientist Take Home Test

Dataset Overview: The data contains information on withdrawals and deposits for a large bank in the Middle East. The original dataset has information for 1100 branches spread over 15 years, however for the sake of brevity we are restricting ourselves to daily aggregated data for just 4 branches over 1 year.

Objective: For small transactions (typically $1000 or less), people go withdraw money from ATMs. For large transactions (say $20,000 or so), people need to visit their branch and withdraw money from there. Consequently, every branch keeps millions of dollars in their vault at any given time to meet customer requirements.

The problem with keeping so much extra money is that a large amount of this cash lies in the vault, totally unused, and thus doesn't earning interest. If the bank knew exactly how much money would be withdrawn in any given day for each branch, they would be able to keep just the right amounts of cash and earn interest by lending out the remaining cash which is no longer needed. Optimizing money this way results in huge annual savings for the bank, while still meeting customer requirements.

Your job is to predict daily cash withdrawals per branch ("Branch Cash Withdrawal" in the dataset), using any machine learning framework (language / statistical model / library) of your choice. Optimize your models to obtain the lowest possible RMSE.

Steps: The below steps are just guidelines, feel free to go back and forth as required.

1. Import data
2. EDA
3. Cleaning and preprocessing
4. Graphs and visualizations
5. Train test split
6. Modeling
7. Hyper parameter tuning, feature engineering, ensemble modeling etc.
8. Final model evaluation
9. Future Predictions: please make daily withdrawal predictions per branch, for Jan 2020.

Note:

1. We would like to see all models you've tried, not just the one with the best error metrics
2. Please comment your work at every stage
3. Please check for errors before submitting
4. Since this is a bank in the Middle East, the weekend falls on Friday / Saturday. All branches are closed on weekends.
5. As this is confidential bank data, we cannot send out more parameters for you to use. This will obviously affect RMSE, but we want you to try and make the best of the

features available. Try and create more columns from the existing columns to improve your model.

6. Remember, not all columns can be used in your forecasts for this information may not be available later. You'll need to ignore these columns, if any.
7. If in doubt regarding what a field stands for, use your best judgement and move ahead.

Output: Please email your work for review by Sunday, 1st March 2020. Do the following:

1. Attach your script for review.
2. Clearly state your best RMSE in the email body.
3. Attach your predictions for daily withdrawals per branch for the month of January 2020 as a CSV file.
4. The subject line of your email should be "Data Scientist Take Home Test".
5. Finally, please send your email to rohit.bhattacharjee@maaloomatiia.com

***