**PROJECT TOPİCS**

**Boğaziçi University**

**Department of Economics**

**Special Topics on Financial Application of Machine Learning**

**Course: EC48E**

Burak Saltoğlu

**Guidelines About the Machine Learning Project**

We expect mainly an applied work using the learning techniques we will use in the class. The research question should be clear. Techniques should be implemented neatly. Pros and cons of each method should be discussed. Your results should be presented in a very organized way. For graduate students I expect a little more since they are more experienced in writing and reading papers. In addition, those who have better coding background may choose more challenging project topics. There are some more theoretical topics that can be chosen by some graduate students.

If you have further question write our TA’s:

İslam Tarlacı : ibahadirtarlaci@gmail.com

**Time line:**

* By 15 of April you should let us know your topic chosen.
* By 15 of May first progress report should be submitted.
* By the final exam date you should hand in your project.
* We may charge 5% daily penalty for the late submission.

**Presentantion**

* you should submit your work in both R Mark Down and a pdf file.
* All the results/tables/graphs’ models data should be reproducible via R Mark Down

**The project presentation format should follow an academic paper format**

1. Abstract, (short summary your project. Please do it with only your own wording. There are very many different web sources to correct your English and grammar.)
2. Introduction, (1 or 2 pages introducing the topic your ideas what you want to do in the project, why is it relevant etc)
3. Literature Review, (very briefly review the relevant literature in the field)
4. Methods, (which methods you have used and justification of your choice in a technical way)
5. Data, (explain the source, reliability and relevance of data, with some nice visualization techniques you learnt in the course)
6. We do not accept Kaggle or similar source of data.
7. Empirical Results (Please do not produce very many tables results without any structure. You do not have enough space to put pout what you found. Be very clear what to present and how to organize your presentation. Do only show th most important findings. Part of the ML projects is to sell your results. People do not have infinite time and patience to follow your endless graphs or solutions even if they may be very useful. Don’t copy and paste your results please present in a very clear and understandable way)
8. Comments on your results (most important part: be very clear what you found. What did you find? Why is it relevant? Where can we use you model academia or industry? Don’t exaggerate your findings but do not underestimate if you feel it is different than what people found previously. Write an honest section to tell what are your own findings with your own words).
9. Conclusion. (summarize your project within half a page. Tell the reader what else would you do if you had time or given an opportunity to write another project).

Notes:

1. You are not allowed to use a topic which you submitted for another course. You may use some follow up studies from your other courses but please inform our course TA or me beforehand. We may or may not approve your topic depending on your potential contribution in our course. Read what other people wrote.2
2. Please inform us on your topic in our course. We do not too many groups making project on the same topic.

**Further Comments and Suggestions:**

For experienced reviewers like professors or academic referees, discerning a well-organized and well-written project from an ordinary one is relatively simple. While obtaining a dataset and estimating machine learning models might be straightforward, the key differentiator lies in the depth and quality of your analysis and project organization. While, insightful analysis is crucial, presenting your findings meticulously is equally important. A polished presentation can elevate even the most brilliant work, while a sloppy one can bury even the deepest insights. This means dedicating significant time and effort to organization, not just analysis.

Projects completed in the rush towards a deadline often suffer from disorganization, lack of clarity, and poor presentation. To avoid these pitfalls, it's crucial to engage deeply with your project from the outset. Instead of racing against time, focus on enriching your understanding through extensive reading, thoughtful analysis, and careful organization. Pay attention to the structure of well-written papers to learn how effective organization can enhance your work. Allocate at least 30% of your time to meticulously organizing your paper to ensure coherence and impact.

The first step is choosing a topic that genuinely interests you. This genuine interest will be the driving force behind your commitment during the analysis and organization phases of your project. However, to ensure originality, avoid topics recently covered in another course. While incorporating follow-up studies from past courses is acceptable, please keep your TA or me informed beforehand.

Ensure effective teamwork by dividing tasks appropriately among team members. Our policy assigns three individuals to each project, as we believe collaborative efforts yield better results than solo endeavors. If you're not well-acquainted with your classmates, we will help by pairing you with teammates. View this as a valuable chance to hone your teamwork skills, which are essential for success in both academic and professional environments.

**Evaluation**:

The quality will be determined on the format, organization, originality of the project. Putting it in the right context with the right literature review makes it more valuable. It should be *well organized*, should be following the following structurally development phases

1. Abstract
2. Introduction
3. Literature Review
4. Data and the Empirical Results
5. Commenting your Finding
6. Conclusion

Your project should be very easily followed so any reader should be able to understand. Projects produced in short span of time usually are lacking good organizational attributes. A good project should make the reader to get engaged and read it very carefully. A sloppy application of some R or Pyhton ML libraries with less emphasis on results obtained will a make a project to look very ordinary or weak. Please do not use a method that we did not cover in the class. If you want to add something different you have to get permission from our TA. We will noy evaluate the models we did not cover in our class. Your R or Python Codes used in the project should be written neatly and re-producible.

*Again: Do not use your previous projects developed for your other courses.*

Towards to end of term we may share some of our older projects submitted for the course.

TOPICS TO BE CHOSEN

1. *Classification of high vs low volatility regimes of USDTRY (or any other currencies) using machine learning*

Financial literature is full of evidence of time varying volatility in asset returns. In some days we observe calm markets with low volatility, and in some other days we observe huge fluctuations with high volatility. USDTRY is a good example to observe this. In this project, the main question is predicting whether we will have a high or low volatility during the next month. For this, collect USDTRY data from CBRT and calculate daily returns. For each month calculate within month volatility using daily returns.:

Our aim is to classify the next month (i.e. Low Volatility Month or High Volatility Month) by using past data. Use time series cross validation in your modelling, so you cannot see the future to predict the past. For the training period, identify low and high volatility months using one of these: a) a volatility threshold above which the month is categorized as a high volatility month (say 20% and above). For the test period, classify the next month using any classification algo from the machine learning literature. H. ere you can use past returns of USDTRY to calculate different features (e.g. past volatility, past skewness and kurtosis, past trends, max/min returns, return streaks, etc), plus you can generate features from other macro and financial variables (e.g. VIX index, BIST100 index, Central Bank interest rates, etc). Feature engineering is a critical step for the success of your model. Thus in this step consider the stationarity of your features and model accordingly. Once you have predictions, evaluate the performance of your predictions using a confusion matrix and other performance statistics. Before performing this last step, write down your own subjective view about which features you think are the most important ones in predicting USDTRY volatility. Then perform a feature importance study and identify the most important features from the ML model. Compare the results with your own subjective views. Comment whether are there any areas that the ML algorithm and/or you as a human can learn from each other's findings and how to improve the model if you will have time after taking this lecture .

1. *Modeling Growth Dynamics Through ML methods*

Same as above but study Turkish quarterly growth rate rather than exchange rate volatility. In this project, good years and bust versus bad years and recessions or crisis. What are the features to be used in classifying good vs bad growth performance? Divide the data into test and training data. Evaluate your growth prediction model.

1. This same method can be applicable for monthly inflation series. 2002-2020 data would be sufficient. Use some international data such as global food prices, exchange rate changes etc. Find the main features to explain high and low inflationary regimes. Similar steps in the above method can be used.
2. *Yield Curve Modeling via ML methods*

Is the “yield curve slope” of US economy a good predictor or GDP recession? There is a large literature on this. (one general source <https://www.bis.org/publ/confp02n.pdf>). First collect economically relevant variables to predict recessions. Then by using the Machine Learning techniques we used try to reduce the dimension and use slope of yield curve. we have studied in the course. Data is generally available in <https://fred.stlouisfed.org/>.

1. Same as above question but use the Turkish TL yield curve data. We can help you to obtain the Turkish data.
2. *Macro Stress Testing via ML Methods*

What are the macroeconomic determinants of the credit defaults in the Turkish economy. Use supervised learning methods we used in the class to study the relationship of GDP growth rate, Unemployment rate, Loan Growth rate, interest rates,VIX,…. Use the quarterly nonperforming loans for the whole banking system given in CBRT. We can discuss about the data set once you have decided. Some papers and ideas: <https://www.researchgate.net/publication/256055499_A_Framework_for_Macroprudential_Bank_Solvency_Stress_Testing_Application_to_S-25_and_Other_G-20_Country_FSAPs>

1. *Exchange Rate Modeling with ML*

How to forecast exchange rates via Machine Learning techniques. Use $/TL or €/$ etc? It is a regression analysis with time series data.

1. Sociology or Political Science: Collect some of the older polls results on election prediction. Use forecast combinations methods to improve on the poll companies. Discuss how can you improve these poll forecasts by using “ensemble methods” in machine learning.
2. Can we predict crude oil price volatility via some sentiment data? Use geopolitical sentiment data, economic uncertainty indices and political uncertainty indices to predict crude oil volatility.
3. Some currencies such as Krone of Norway and Canadian Dollar are known as Commodity currencies meaning their value very much depend on commodity prices particularly crude oil prices. Can you develop an ML model for these two currencies to test whether the commodity prices and these exchange rate are moving together? Can you develop an algorithmic trade on the crude oil prices and CAD or Norwegian Krone? If your model produce good enough predictions you can trade these currencies as commodity prices change. When crude oil prices go up or down you make profit in buying and selling these currencies.

More conceptual (MA/PhD Students): discuss the advantages of various prediction performance measures such as cross validation etc. Discuss various forms of cross validation tools when you face with time series data. What are the critical aspects of model selection in time series data?

1. *Casuality and ML Methodology*

More theoretical and advanced (MA/PhD students): Machine Learning methods have been criticized by ignoring the “causality” among variables. Some recent papers https://www.jmlr.org/papers/volume16/bontempi15a/bontempi15a.pdf [https://www.nature.com/articles/s42256-020-0197-y.. It](https://www.nature.com/articles/s42256-020-0197-y..%20%20It) is also said that ML 2.0 will be based on models integrating causality into the conventional methods. Write a very clear essay discussing the present and the expected ML models in the near future.

1. *Discuss the recent development in Gradient descent model. There are variants of Gradients Descent Batch Methods, Mini Batch GD, Stochastic Gradient Descent among many others. In a concrete optimization problem chosen from an ML model show the advantages and disadvantages of each of these models. The speed, accuracy, convergence rates etc. (Math or Engineering Students might also choose this topic)*
2. *Data Scrapping and Sentiment Analysis with Alternative Data*
3. *Option Implied Volatility and Machine Learning*

What is the best model for foreign exchange volatility for Dollar/TL. Compare, econometric models and option implied volatility models. Can we use forecast combinations in ML to predict better volatility. Classical paper by Jorion: [https://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.1995.tb04793.x ..Many more on..will](https://onlinelibrary.wiley.com/doi/10.1111/j.1540-6261.1995.tb04793.x%20..Many%20more%20on..will) be helping in finding data.

1. *Special Topics*:

If you have a different but solid idea similar to the above topics we may allow to work on a well defined topic. Clearly state your project topic in a written form so that we can have a look and decide if you can proceed with this topic. Then you may do your own topic.