**Final Exam** **Stevens Institute of Technology**
**Professor Germán G. Creamer, Fall 2020**

I pledge on my honor that I have not given or received any unauthorized assistance on this examination. I further pledge that I have not copied any material from a book, article, the Internet or any other source according to the exam's instructions.

**Signature:** _____ **Name:** _____ **Date:**_____

Answer all the following questions and justify your answer when required. The exam is an open book exam. You should neither collaborate with anybody nor search online solutions to the exam's questions. However, you can access the link included in this document or access financial websites that provide information of the stock under study. Total: 25 points. 1st part: questions: 1.5 points each question; 2nd part: 13 points. (if the Python code file is missing, 2.5 points are deducted).

You are part of a research team of a FinTech company dedicated to forecast the price direction of public US companies. Using the attached dataset, select the company assigned to you according to your Stevens ID in the tab "assignments."

You must build a forecasting model of your assigned company's next day's return trend using the years 2014-2018 to train and 2019 to test your model.

The dataset is based on the Fama French 3 factor model:

$$r_{i,t} - rf_t = \alpha_i + \beta_{i,mkt} mktrf_t + \beta_{i,smb} SMB_t + \beta_{i,hml} HML_t + \varepsilon_{i,t}$$

Where

- $r_{i,t}$= stock $i$ return during period $t$
- $mktrf_t$= Fama French Excess Rerturn on the Market during period $t$
- $SMB_t$= Fama French Small Minus Big (Size) factor during period $t$
- $HML_t$= Fama French High Minus Low (Value) factor during period $t$

The dataset includes the following variables or columns based on the calibration of the above model using N (252) days:

Ticker: identifier of the stock
Permno: permanent number of the stock
Date: date of observation
N: number of observations used to compute Beta
Ret: daily stock return i
Alpha: Alpha of stock i
b_mkt: Beta on $mktrf_t$
b_smb: Beta on $SMB_t$
b_hml: Beta on $HML_t$
ivol: Idiosyncratic Volatility: Variance of residuals or the part that cannot be explained by the Fama French model
tvol: Total Volatility: Variance of stock returns

R2: R-Squared. Fitting measure of the Fama French equation for date t
Exret: Excess Return from Risk Model: stock return I (Ret) – risk free rate

Further details and data of the Fama French model can be obtained at:

http://mba.tuck.dartmouth.edu/pages/faculty/ken.french/Data_Library/f-f_factors.html

https://mba.tuck.dartmouth.edu/pages/faculty/ken.french/data_library.html

**Questions:**

**1st. Part: Please answer the following questions justifying your answer. You SHOULD NOT INCLUDE any results or Python code in this section. This section only helps you to propose your solution.**

1. Include a brief description of the stock under study and possible factors that may affect its performance.
2. What precisely is the forecasting problem? How do you plan to calculate it? What is the target variable in your machine learning algorithm?
3. Which three forecasting machine learning algorithms do you think are appropriate for this problem domain and why? How do you plan to select the best algorithm? Justify your answer.
4. What features would you use? You do not have to use all the variables included in the dataset, and you could also use additional variables. Would you modify or correct your data in any way? If you plan to introduce changes to any variable, you must be very specific and tell what variable(s) will change and how.
5. How will you calibrate your model and evaluate whether your model has captured any generalizable knowledge? Explain your method, and justify the metric(s) that you propose to employ.
6. Compare two graphs that you could use to evaluate the different algorithms' performance and explain how they can help evaluate your models and select the best method. Select one graph and justify your answer.
7. How can you rank the features' importance?
8. Once you put your system in production, would you leave it to work alone, or what else can you do?

**2nd. Part:**

Based on your answers to the first part of the exam, implement your solution using Python. Make sure that you cover all the steps and explanations to questions 2-7.

1. Present and DISCUSS the results of your Python program, answering the above questions.
2. Write a conclusive paragraph about the quality of the forecast of your stock. Indicate if you suggest to your client buy, sell, or do not trade the stock under evaluation for the first day of 2020 (you should not use the historical information of the stock in 2020 to answer this final question).

**Report and submission:**

You should prepare your final exam report as a Jupyter notebook. You can answer the first part as a large text section. The second part should have numbers that correspond to questions 2-7 of the first part. Explain your results or the code of each section. You should submit two files: 1). a Jupyter notebook as an HTML or PDF file, and 2). a file with your Python code (.py extension).