

Time Series Forecasting in Python

Introduction

Time Series Forecasting is something of a dark horse in the field of data science. It is one of the most applied data science techniques in business - used extensively in finance, in supply chain management, and in production and inventory planning. Moreover, it has a well established theoretical grounding in statistics and dynamic systems theory. Yet, it retains something of an outsider status in data science compared to more recent and popular machine learning methods such as image recognition and natural language processing. Consequently, Time Series Forecasting gets little or no treatment at all in introductory data science and machine learning courses.

This course is intended to provide a comprehensive introduction to forecasting methods without deep diving into the theoretical details behind each method. Although, the references at the end of each week will fill in many of those details.

The course is intended for the following three audiences.

1. Graduate or PhD students studying in STEM or business fields.
2. People doing forecasting in business who may not have had any formal training in the area.
3. MBA students doing a data elective.
4. Also relevant for those studying public policy, healthcare management, and related disciplines.

Course Relevance

- Time Series Forecasting is one of the most applied data science techniques in business - used extensively in finance, in supply chain management, and in production and inventory planning.
- If you are a data professional, there are very high chances that you will have to analyze time series data in your career.
- Used across all sectors and industries.

Prerequisites

- It is assumed that course participants are familiar with

- Introductory statistics.
- High-school algebra.
- Graphical methods for describing quantitative data - line plots, histograms, barplots, box plots, correlation plots.
- Numerical measures of Central Tendency.
- Numerical measures of Variability.
- Using the Mean and Standard Deviation to describe data.
- Joining data - inner, outer, right, and left join.
- Hypothesis testing.
- Simple linear and multiple regression - model assumptions, fitting the model, assessing utility of model.
- Splitting data into test-train splits and cross validation.
- Model performance evaluation metrics - RMSE, MAE, MAPE.
- Basic Python knowledge required - ability to read data and ability to perform basic manipulations using pandas and numpy libraries.
- If you don't have prior experience with the aforementioned, you can still enroll if you have the willingness to learn.
- Bring your enthusiasm and curiosity.

Technology Stack

We will use Python as a tool to analyze time series data and generate forecasts. The intention is for students to learn how to manipulate and wrangle data with Python. Python is free and available on almost every operating system. It is a wonderful tool for all statistical analyses, not just for forecasting. Also, it is one of the most widely used programming languages in business. Please have **Jupyter Notebook** installed on your device as we will be using it for data analysis.

Learning Objective

- Forecaster's Toolbox - Build and apply time series forecasting models in a variety of business contexts using tools that are useful for many different forecasting situations such as the following.
- Benchmark/Simple forecasting methods.
- Ways of making the forecasting task simpler using transformations and adjustments.
- Residual analysis - methods for checking whether a forecasting method has adequately utilized the available information.
- Techniques for computing prediction intervals.

- Be able to use a range of forecasting methods - ETS (Error, Trend, Seasonality), Holt Winter's, ARIMA (Autoregressive, Integrated, Moving Average), SARIMA, VAR (Vector Autoregression), and a suite of machine learning models such as XGBoost, Random Forest, Support Vector Regression, etc.
- Be able to use graphical methods to explore the time series data, analyze the validity of the models fitted, and present the forecasting results.
- Develop skills, mindsets, and behaviors that are most sought after in the industry today - data science, product management, entrepreneurship, and storytelling.
- Don't think like a Data Scientist / Software Engineer / Solutions Architect / Tech Consultant. Think like a CEO. Take an end-to-end view.

Learning Process and Course Format

- This course is mostly 100% Virtual and runs synchronous. I live in San Francisco and sometimes travel to campus in Pittsburgh, PA. There is an opportunity to meet face-to-face at such times.
- All lectures are scheduled via Zoom on Wednesdays, 6.30-9.20 pm EST. The links will be provided on Canvas. Go to the zoom tab on the left-hand side. **All lectures will be recorded.** I'll be sharing the recordings after class via email. Please make sure that your Internet connection and equipment are set up to use Zoom. Let me know if there is a gap in your technology set-up as soon as possible, and we will work together to find a solution.
- **Attendance Policy** - mandatory - 5% of the grades are towards attendance and engagement in class. If you are consistently unable to attend the live zoom sessions due to any reason, you may stay registered and watch the lecture recordings afterwards but will lose the 5% participation points for not attending live.
- **For the purpose of building a classroom community to drive better learning outcomes, it is mandatory to turn on your cameras during class.** Having a virtual background is completely fine. We are already physically distanced. Let's not be physically invisible too. If you cannot share your video for some valid reason, please let me know beforehand. Some or all class engagement points will be deducted for not turning on your cameras without giving me a heads up on a consistent basis.
- **Assignments** - This course has one quiz and two assignments that you need to complete individually. We do not have a final exam but we have a

final group project that you need to complete working in groups with your peers. I'll be sharing more details in class. All expectations and submission deadlines will be shared in class.

- Class presentations, datasets that we will be exploring, supervised readings, and any other content to supplement the learnings will be shared in class and via emails.
- How to reach out to me? This course does not have a TA. If you have any questions about the coursework, submissions, or any other questions in general, please feel free to email me (tshaikh@alumni.cmu.edu) or message me on Canvas and I will get back to you as soon as I can. Generally, I get back within 2-3 hours.
- Want to learn about me? Here is my LinkedIn profile. Hit me up.
<https://www.linkedin.com/in/tanvirahmedshaikh/>

Key Topics that we will cover in this course

1. Univariate and Multivariate Time Series Forecasting using techniques such as ETS, ARIMA, SARIMA, and VAR models in addition to machine learning models such as XGBoost, Random Forest, SVR, etc
2. Product Strategy and AI product Management
3. Storytelling with Data
4. Python for Data Science
5. Influence decision makers (influence without authority)

Learning Resources

1. Textbook recommendation (father of forecasting Rob J Hyndman).

Forecasting: Principles and Practice - <https://otexts.com/fpp2/>

2. Free online course recommendation -
<https://www.udacity.com/course/time-series-forecasting--ud980>

Tentative Course Schedule -

Weekly 3 hour lecture (online) having the following content.

Week 1

1. History of forecasting
2. What is Time Series Forecasting? Examples.
3. Why is forecasting important in business?

4. What can be forecasted?
5. Types of forecasts
6. Forecasting concerns
7. Basic Steps in Forecasting
8. Why use Python?
9. Lab: Set Up Jupyter Notebook/Lab Environment

Week 2 - Hands On Session - Predict median monthly stock price

1. Install and import packages
2. Download single and multiple stock data - explore joins
3. Dataframe in Python - it is a TABLE!
4. Explore the data using functions in Python
5. Create visualizations using functions in Python
6. Forecast the median monthly stock price
7. Evaluate models
8. Save the downloaded data to - csv and excel
9. Assignment 1 released

Week 3 - Predict monthly anti-diabetic drug sales in Australia

1. White Noise
2. Autocorrelation
3. Stationarity - what is it and how to identify it?
4. Differencing
5. Autoregressive models
6. Moving average models - not like your normal MA
7. ARIMA - seasonal and non-seasonal
8. Quiz and Assignment 2

Week 4 and 5 - Multivariate Multi Output Time Series Forecasting- Predict Monthly Average Ozone Air Quality Index in San Francisco

1. Dataset used - US Air Pollution data
2. Statistical Multivariate Models - VAR and SARIMAX
3. Multivariate ML Models - we will answer the following two questions
4. What are the leading indicators that explain what the average Ozone Air Quality Index will be over the next 12 months?
5. Predict monthly average Ozone Air Quality Index for each of the next 12 months.

6. ML Models - Feature Engineering / Add useful predictors and Variable Importance
7. ML Multi Output Models - Convert time series data to multi step or multi output supervised learning problem
8. Time Series Cross Validation and model selection
9. Assignment 3 - Final Assignment
10. Break into teams to solve a real world forecasting problem

Week 6 - Storytelling with Data and Product Strategy

1. Types of analysis - exploratory and explanatory
2. Examples of good and not so good visualizations
3. How to communicate visually with data?
4. How to tell a story - Six Essential Elements of a Story
5. Example - Work Presentation
6. Product Strategy Framework - V2MOM
7. Group Assignment Refresher

Week 7 - Gallery Walk!

1. Present your findings. Team presentations.
2. Tools of an entrepreneur - Jobs-to-be-Done, Discovery Hypothesis, Buyer Persona, Business Model Canvas, Value Proposition Canvas
3. Life advice - Steps to Build Wealth and Book Recommendations

Course Assessment

The final course grade will be calculated using the following categories.

Assessment	% of Final Grade
Stationarity Quiz	10%
Assignment 1	25%
Assignment 2	25%

Group Project	35%
Attendance + Class Participation	5%

Grading Scale

Even though grades really don't matter in life, students will be assigned the following final letter grades, based on calculations coming from the course assessment section.

Grade	Percentage Interval
A+	97% and above
A	93-96.99%
A-	90-92.99%
B+	87-89.99%
B	83-86.99%
B-	80-82.99%
You won't get below this.	