Data Analytics

Lesson 09.

Investigating Predictive Analytic Techniques

Dr. Hai Tran

hai.tran@sbsuni.edu.vn

Scholar: https://scholar.google.com/citations?user=kHZvlTkAAAAJ&hl=en&oi=ao

Co-Founder: XAI - https://xai.foo/

SBS | Saigon Business School

In partnership with

UCW UNIVERSITY CANADA WEST

MACQUARIE University SYDNEY·AUSTRALIA

UO N University of Northampton

AMITY GLOBAL INSTITUTE

Learning materials

● Textbook
  ● Evans, J. (2016) Business Analytics. 2nd edn. Pearson.
  ● Runkler, T. (2016) Data Analytics: Models and Algorithms for Intelligent Data Analysis. 2nd edn. Vieweg+Teubner Verlag.

● Online reference materials
  ● archive.ics.uci.edu/ml/
  ● powerbi.microsoft.com
  ● https://github.com/topics/data-analysis-python
  ● https://media.pearsoncmg.com/ph/esm/esm_evans_eba3e_20/tools/eba3e_analytic_solver.html
  ● https://data.imf.org/

Agenda

- Lesson 1: Understanding Data Analytics Terminologies.
- Lesson 2: Foundation of Business Analytics
- Lesson 3: Visualizing and Exploring data
- Lesson 4: Applying Descriptive Analytic Techniques
- Lesson 5: Data Modeling
- Lesson 6: Predictive Analytics
- Lesson 7: Regression, Classification and Clustering
- Lesson 8: Forecasting Techniques
- Lesson 9: Investigating Predictive Analytic Techniques
- Lesson 10: Introduction to Data Mining
- Lesson 11: Demonstrating Prescriptive Analytic Methods
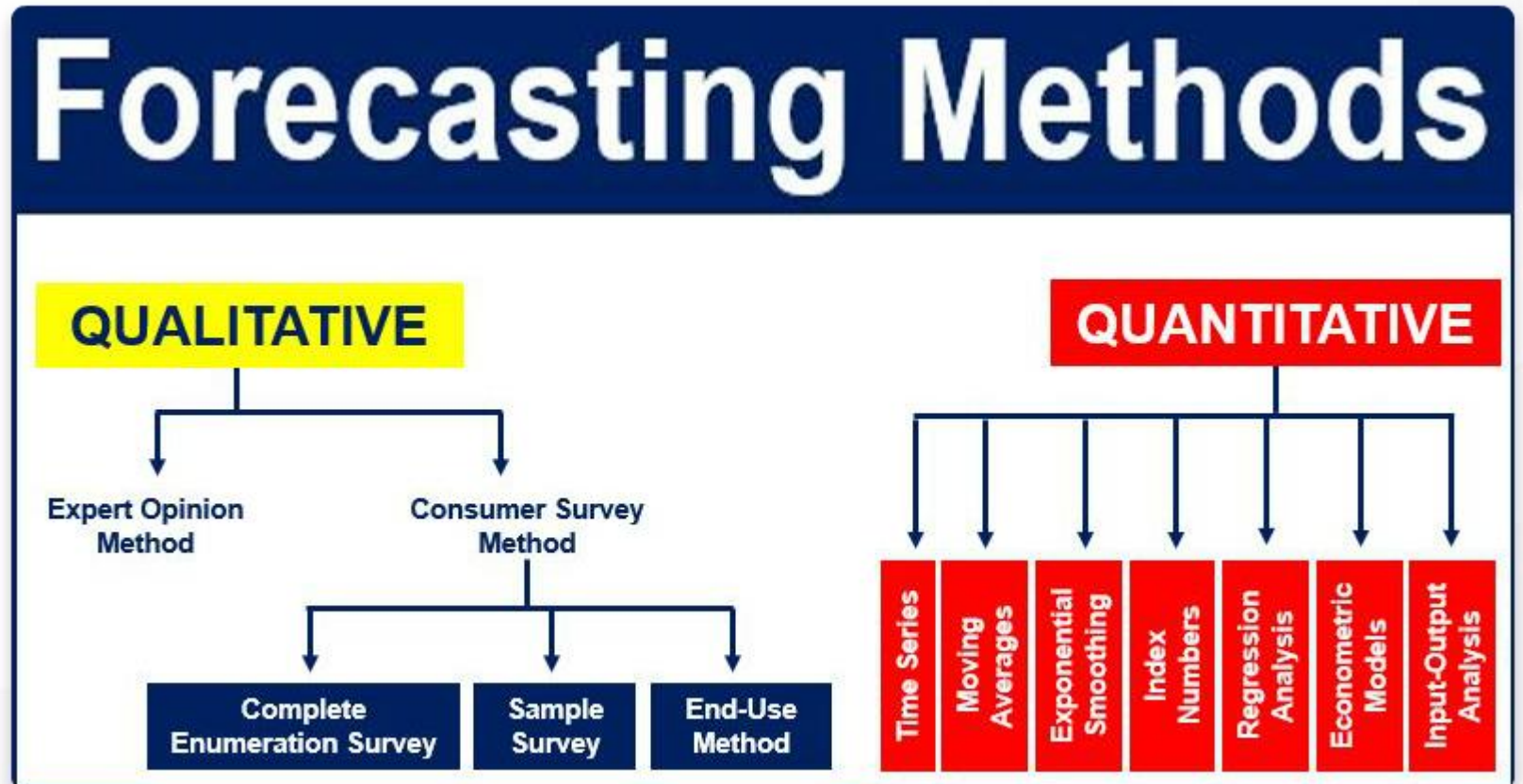- Lesson 12: Recap and advanced topics

Two main forecasting approaches are qualitative and quantitative.

The **qualitative** method is based on expert opinions and the comprehensive analytical research of consumers' behavior.

The **quantitative** method is built on the concept of past statistics research.

Discover the importance of forecasting and learn about the different types of forecasting techniques in this engaging presentation.

Forecasting Methods

QUALITATIVE

- Expert Opinion Method
- Consumer Survey Method
  - Complete Enumeration Survey
  - Sample Survey
  - End-Use Method

QUANTITATIVE

- Time Series
- Moving Averages
- Exponential Smoothing
- Index Numbers
- Regression Analysis
- Econometric Models
- Input-Output Analysis

Investigating Predictive Analytic Techniques

```python
from scipy import stats

x = [5,7,8,7,2,17,2,9,4,11,12,9,6]
y = [99,86,87,88,111,86,103,87,94,78,77,85,86]

slope, intercept, r, p, std_err = stats.linregress(x, y)

def myfunc(x):
  return slope * x + intercept

speed = myfunc(10)

print(speed)
```
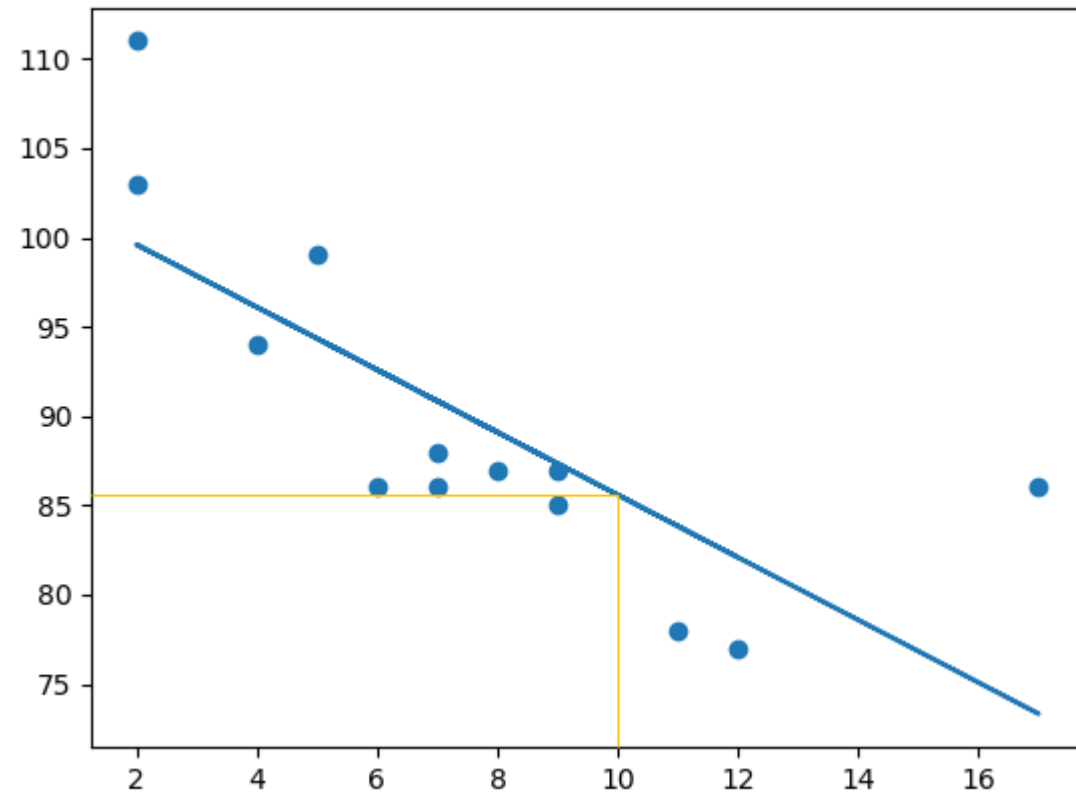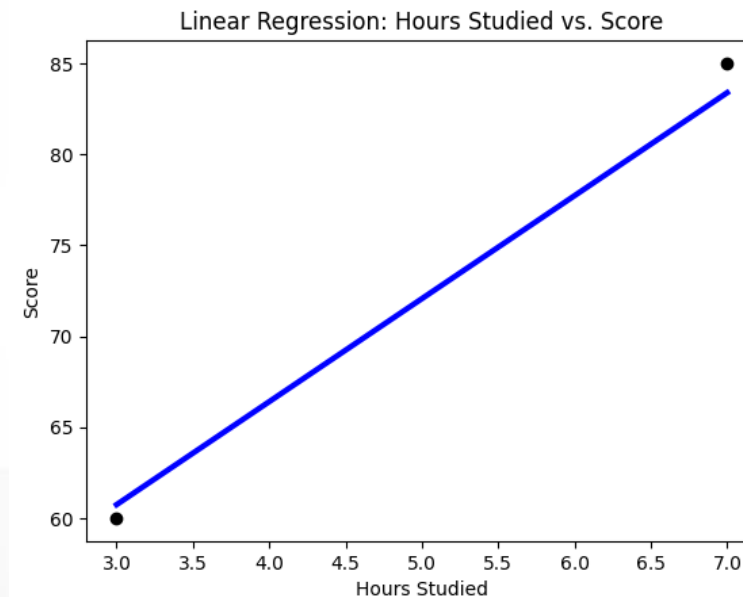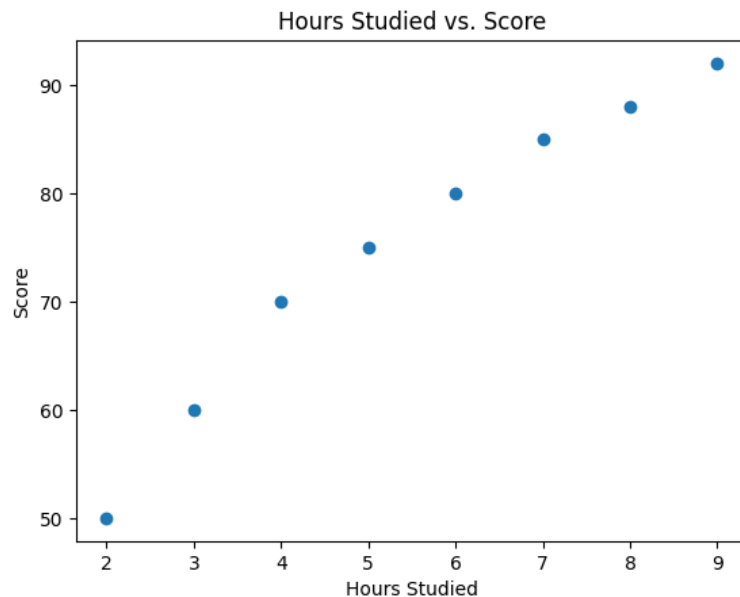
- Predictive analytics using regression is a common task in data analytics. Let's consider a example where we use linear regression to predict the **score** of students based on the number of ***hours*** they ***study***.

- ```
  # Create a simple dataset (you can replace this with your own dataset)
  ```

- ```
  data = {'Hours': [2, 3, 4, 5, 6, 7, 8, 9],
  ```

- ```
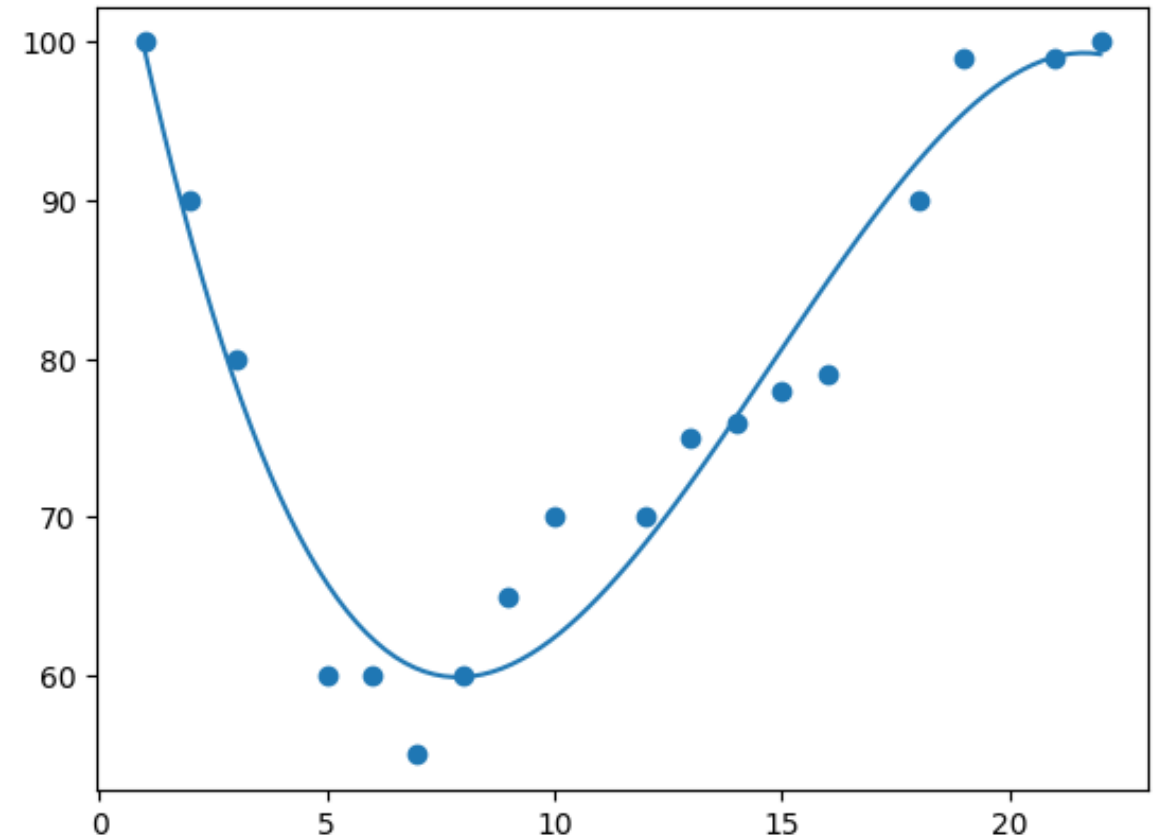   'Score': [50, 60, 70, 75, 80, 85, 88, 92]}
  ```



sbsedu.vn

```python
import numpy
import matplotlib.pyplot as plt

x =
[1,2,3,5,6,7,8,9,10,12,13,14,15,16,18,19,
21,22]
y =
[100,90,80,60,60,55,60,65,70,70,75,76,78,
79,90,99,99,100]

mymodel = numpy.poly1d(numpy.polyfit(x,
y, 3))

myline = numpy.linspace(1, 22, 100)

plt.scatter(x, y)
plt.plot(myline, mymodel(myline))
plt.show()
```

- To demonstrate forecasting survival in the Titanic dataset using Python, you can use a machine learning approach, such as logistic regression. Below is a simple example using Python with the popular scikit-learn library:

```python
# Import necessary libraries
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LogisticRegression
from sklearn.metrics import accuracy_score,
classification_report
from sklearn.preprocessing import LabelEncoder

# Load the Titanic dataset
titanic_data = pd.read_csv('titanic.csv')

# Preprocess the data
# Drop unnecessary columns or fill missing values as
needed
titanic_data = titanic_data[['Pclass', 'Sex', 'Age',
'SibSp', 'Parch', 'Fare', 'Embarked', 'Survived']]
titanic_data = titanic_data.dropna()

# Encode categorical variables
label_encoder = LabelEncoder()
titanic_data['Sex'] =
label_encoder.fit_transform(titanic_data['Sex'])
titanic_data['Embarked'] =
```

- To predict a new record, you can use the trained model to make predictions on a set of features for a new passenger. Here's how you can modify the code to predict a new record:

```python
# ... (Previous code remains unchanged)

# Suppose you have a new passenger's information
new_passenger = {'Pclass': 3, 'Sex': '0', 'Age': 25,
'SibSp': 1, 'Parch': 0, 'Fare': 7.5, 'Embarked': '0'}

# Convert the new passenger's information into a DataFrame
new_passenger_df = pd.DataFrame([new_passenger])

# Encode categorical variables

# Make predictions for the new passenger
new_passenger_prediction = model.predict(new_passenger_df)

# Print the prediction
if new_passenger_prediction[0] == 1:
    print("The new passenger is predicted to survive.")
else:
    print("The new passenger is predicted not to survive.")
```

Google classroom task

- **What is Forecasting? | Process & Benefits of Forecasting**
- https://www.youtube.com/watch?v=M8Kiwv9gDJU
  - Watch and investigate.
  - Submit your answer:
    - YouTube Link
    - Draw a diagram to show Forecasting Process.
    - List down benefits of forecasting.

sbsedu.vn

Investigating Predictive Analytic Techniques

Question	Match A	Match B
1. What is the primary purpose of data cleansing?	A. Identifying patterns in data	B. Removing errors and inconsistencies
2. Which statistical measure represents central tendency?	A. Mean	B. Standard Deviation
3. What is the main goal of exploratory data analysis (EDA)?	A. Summarizing data	B. Discovering patterns and trends
4. In machine learning, what does "overfitting" refer to?	A. Model fitting the training data	B. Model performing poorly on new data
5. Which type of chart is suitable for displaying the distribution of a continuous variable?	A. Histogram	B. Pie Chart

sbsedu.vn

Investigating Predictive Analytic Techniques

Question	Match A	Match B
1. What is the purpose of time series analysis in forecasting?	A. Identifying trends and patterns over time	B. Classifying data into categories
2. Which forecasting method involves using past values to predict future values?	A. Moving Averages	B. Exponential Smoothing
3. What is the primary advantage of using quantitative forecasting methods?	A. Reliance on historical data	B. Subjectivity in decision-making
4. What does the term "seasonality" refer to in forecasting?	A. Regular patterns that repeat at known intervals	B. Random fluctuations in data
5. Which technique involves combining forecasts from multiple sources or models?	A. Ensemble Forecasting	B. Regression Analysis

Investigating Predictive Analytic Techniques

Our example person has registered every tin there was a comedy show in town, ar registered some information about th comedian, and also registered if he/she went not.

Age	Experience	Rank	Nationality	Go
36	10	9	UK	NO
42	12	4	USA	NO
23	4	6	N	NO
52	4	4	USA	NO
43	21	8	USA	YES
44	14	5	UK	NO
66	3	7	N	YES

Rank <= 6.5
gini = 0.497
samples = 13
value = [6, 7]

True / False

gini = 0.0
samples = 5
value = [5, 0]

Nationality <= 0.5
gini = 0.219
samples = 8
value = [1, 7]

Age <= 35.5
gini = 0.375
samples = 4
value = [1, 3]

gini = 0.0
samples = 4
value = [0, 4]

gini = 0.0
samples = 2
value = [0, 2]

Experience <= 9.5
gini = 0.5
samples = 2
value = [1, 1]

gini = 0.0
samples = 1
value = [0, 1]

gini = 0.0
samples = 1
value = [1, 0]
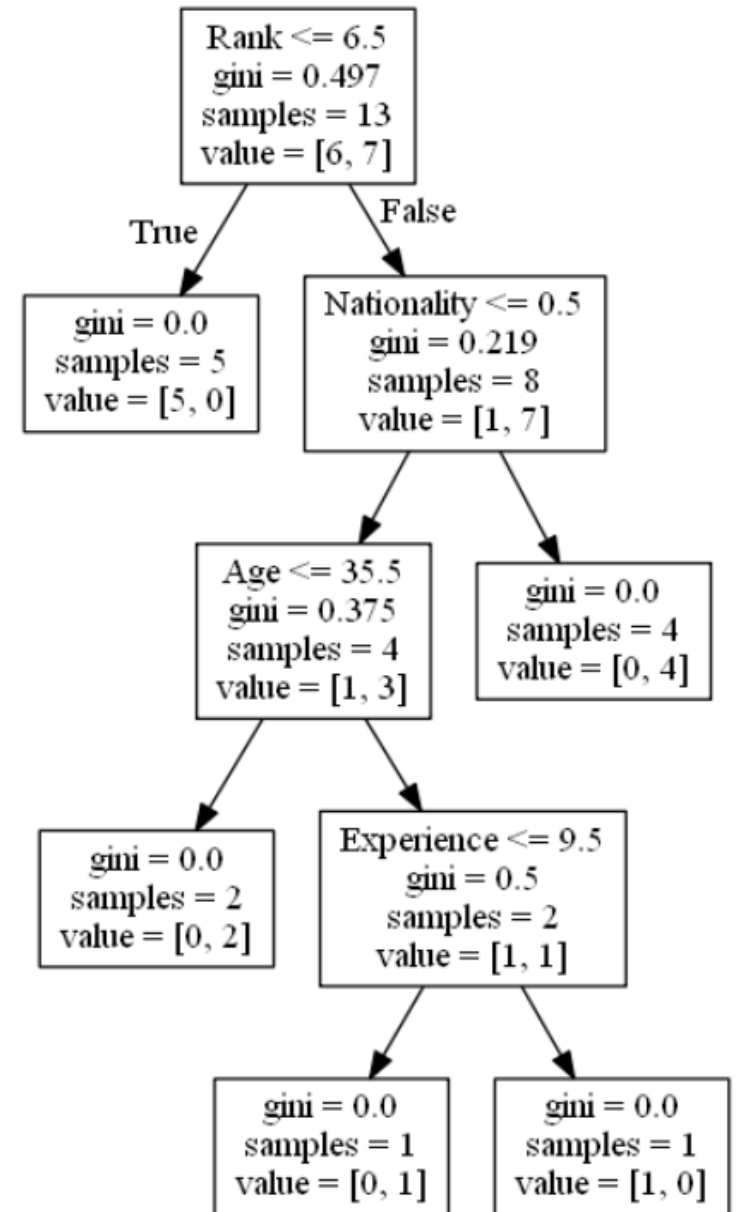
Reading Business Analytics textbook: Chapter 8, and Chapter 9, page 233 – 300.

Discussion and answer:

- What is Mean Squared Error? How to calculate Mean Squared Error?
- What is R-squared? How to calculate R-squared?

$$MSE = \frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2$$

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \hat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \bar{y})^2}$$

Conclusion and Questions

- Future Trends: Get a glimpse into the future of predictive analytics, including AI integration and prescriptive analytics.

- Actionable Insights: Learn how to leverage predictive analytics to make informed decisions and gain a competitive edge.

- Ongoing Learning: Discover resources and training opportunities to continue expanding your knowledge in the field of predictive analytics.

sbsedu.vn

Thank you