PROBABILITY THEORY AND DATA MINING

FIRST STEP TO
BECOMING
A DATA SCIENTIST

INTRODUCTION TO PROBABILITY

- Probability and Statistics
- Basic Concepts
- Rules
- Conditional Probability
- Bayes Theorem

Data Mining Techniques

1. Data Warehousing
2. Data Cleansing and Preparation
3. Association
4. Classification
5. Regression
6. Data Analytics
7. Clustering
8. Artificial Intelligence
9. Machine Learning
10. Association Rule Learning

PROBABILITY DISTRIBUTIONS

- Random Variables
- Bernoulli Distribution
- Binomial Distribution
- Normal Distribution
- Central Limit Theorem
- Mathematical Expectation
- Computer simulation

Data Mining Techniques

1. Data Warehousing
2. Data Cleansing and Preparation
3. Association
4. Classification
5. Regression
6. Data Analytics
7. Clustering
8. Artificial Intelligence
9. Machine Learning
10. Association Rule Learning

STATISTICAL ANALYSIS

- Descriptive Statistics
- Inferential Statistics
- Computer simulation

Data Mining Techniques

1. Data Warehousing
2. Data Cleansing and Preparation
3. Association
4. Classification
5. Regression
6. Data Analytics
7. Clustering
8. Artificial Intelligence
9. Machine Learning
10. Association Rule Learning

IBM SPSS STATISTICS

- Introducing IBM SPSS
- Descriptive Statistics in SPSS
- Computer simulation
- Hypothesis Testing
- Computer Simulation

Data Mining Techniques

1. Data Warehousing
2. Data Cleansing and Preparation
3. Association
4. Classification
5. Regression
6. Data Analytics
7. Clustering
8. Artificial Intelligence
9. Machine Learning
10. Association Rule Learning

INDEPENDENCE

- Concepts
- Computer Simulation
- Odds Ratios
- Chi Square Test
- Fisher Exact Test
- T Independent Test
- Two-Sample T-Test

Data Mining Techniques

1. Data Warehousing
2. Data Cleansing and Preparation
3. Association
4. Classification
5. Regression
6. Data Analytics
7. Clustering
8. Artificial Intelligence
9. Machine Learning
10. Association Rule Learning

VARIANCE ANALYSIS

- Concepts and Implications
- Computer Simulation
- Follow-up Test
- Two-Way ANOVA
- Understanding Covariance Test
- Steps to Implementing Covariance Test

Data Mining Techniques

1. Data Warehousing
2. Data Cleansing and Preparation
3. Association
4. Classification
5. Regression
6. Data Analytics
7. Clustering
8. Artificial Intelligence
9. Machine Learning
10. Association Rule Learning

ROUND UP TOPICS IN STATISTICS

- Correlation Test
- Non-Parametric Test
- AB Test
- Computer Simulation
- Final Statistics Project

Data Mining Techniques

1. Data Warehousing
2. Data Cleansing and Preparation
3. Association
4. Classification
5. Regression
6. Data Analytics
7. Clustering
8. Artificial Intelligence
9. Machine Learning
10. Association Rule Learning

DATA MINING INTRODUCTION

- Introducing Data Mining
- Data Description and Data Mining Methods
- SPSS Modeler Introduction
- Data Entry in SPSS Modeler
- Data Quality
- Handling Out of Range Data
- Handling Outliers and Missing Data

Data Mining Techniques

1. Data Warehousing
2. Data Cleansing and Preparation
3. Association
4. Classification
5. Regression
6. Data Analytics
7. Clustering
8. Artificial Intelligence
9. Machine Learning
10. Association Rule Learning

DATA TRANSFORMATION

- Data Normalisation
- Feature Creation
- Discretisation
- Data Aggregation
- Data Smoothing
- Computer Simulation

Data Mining Techniques

1. Data Warehousing
2. Data Cleansing and Preparation
3. Association
4. Classification
5. Regression
6. Data Analytics
7. Clustering
8. Artificial Intelligence
9. Machine Learning
10. Association Rule Learning

DIMENSIONALITY REDUCTION

- Feature Selection
- Feature Extraction
- Sampling
- Data Integration
- Project: Analysing Customer Behaviour

Data Mining Techniques

1. Data Warehousing
2. Data Cleansing and Preparation
3. Association
4. Classification
5. Regression
6. Data Analytics
7. Clustering
8. Artificial Intelligence
9. Machine Learning
10. Association Rule Learning

RULE BASED PREDICTIVE MODELS

- Introducing Predictive Models
- Decision Tree
- Rule Assessment and Interpretation
- Classification Model Assessment
- Regression Model Assessment

Data Mining Techniques

1. Data Warehousing
2. Data Cleansing and Preparation
3. Association
4. Classification
5. Regression
6. Data Analytics
7. Clustering
8. Artificial Intelligence
9. Machine Learning
10. Association Rule Learning

UNBALANCED DATA

- Challenges in Handling Unbalanced Data
- Implementing Decision Tree
- Confusion Matrix
- Regression Tree in SPSS

Data Mining Techniques

1. Data Warehousing
2. Data Cleansing and Preparation
3. Association
4. Classification
5. Regression
6. Data Analytics
7. Clustering
8. Artificial Intelligence
9. Machine Learning
10. Association Rule Learning

STATISTICAL PREDICTIVE MODELS

- Naïve Bayes
- Linear Regression
- Parameter Estimation
- Model Hypothesis Tests
- Implementing Linear Regression
- Logistic Regression
- Implementing Logistic Regression

Data Mining Techniques

1. Data Warehousing
2. Data Cleansing and Preparation
3. Association
4. Classification
5. Regression
6. Data Analytics
7. Clustering
8. Artificial Intelligence
9. Machine Learning
10. Association Rule Learning

ENSEMBLE PREDICTIVE MODELS

- Introducing
- Stacking
- Bagging
- Boosting
- Implementing Ensemble Learning in Classification
- Implementing Ensemble Learning in Regression

Data Mining Techniques

1. Data Warehousing
2. Data Cleansing and Preparation
3. Association
4. Classification
5. Regression
6. Data Analytics
7. Clustering
8. Artificial Intelligence
9. Machine Learning
10. Association Rule Learning

UNSUPERVISED LEARNING

- Introducing Clustering
- Hierarchical Clustering
- K-Means Algorithm
- DB-SCAN Algorithm
- Association Rules
- Apriori Algorithm

Data Mining Techniques

1. Data Warehousing
2. Data Cleansing and Preparation
3. Association
4. Classification
5. Regression
6. Data Analytics
7. Clustering
8. Artificial Intelligence
9. Machine Learning
10. Association Rule Learning

NOTES

WEBSITES

- Depends On Instructor

PREREQUISITES

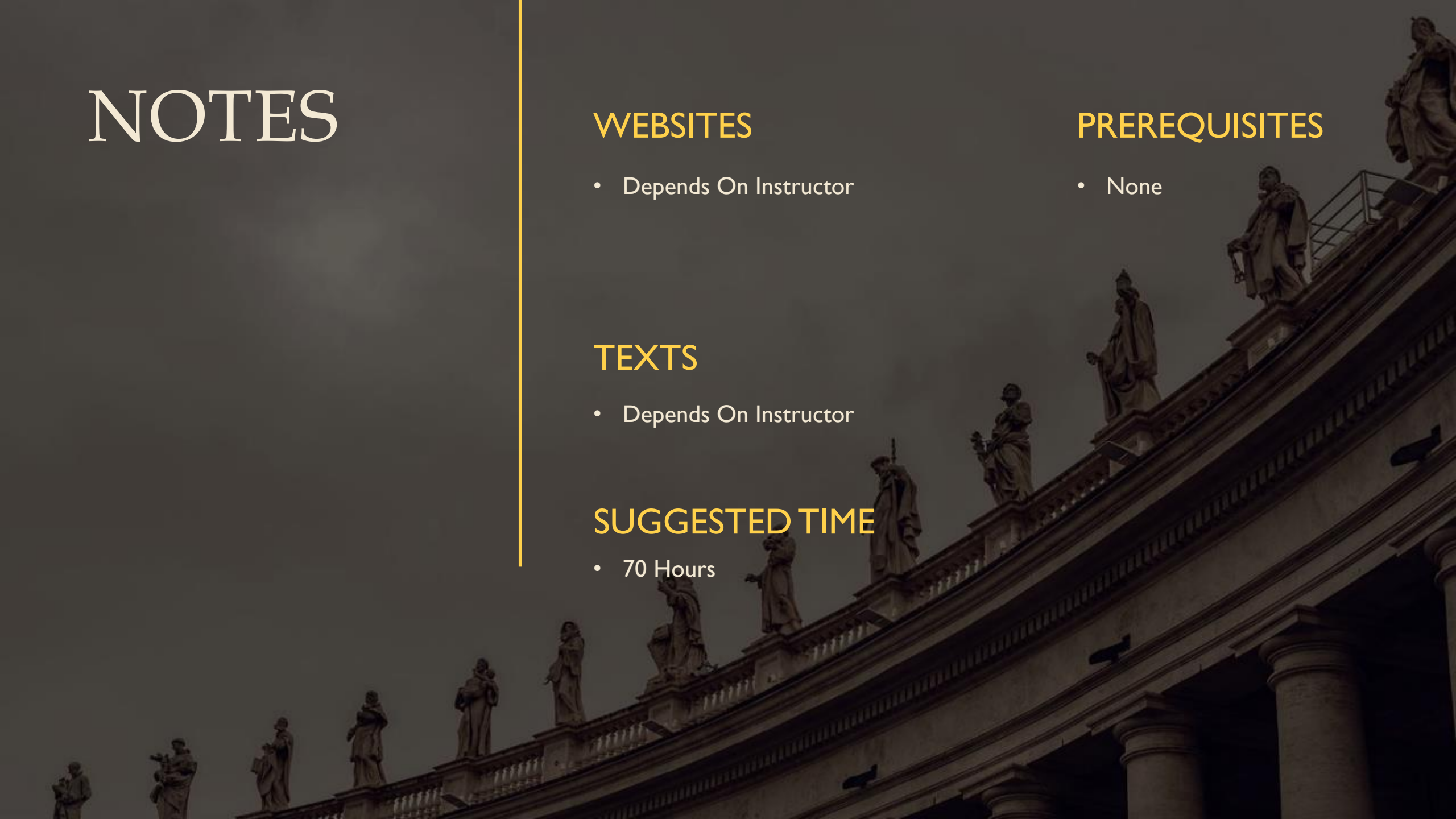
- None

TEXTS

- Depends On Instructor

SUGGESTED TIME

- 70 Hours



DOCUMENT HISTORY

Author	Version	Revision	Date / Time	Department	Validity
Mehdi Shokri	1.0.0		14-05-2023	AI	3 Months