Statistical Analysis – Roadmap to AI&ML

Duration: 40 hours

Overview

This comprehensive course on Algorithms offers a structured progression from foundational concepts to real-world applications, tailoring skills for AI and ML. It begins with

Audience Profile

- AI/ML Enthusiasts & Data Scientists
- Power BI & Data Analytics Professionals
- Software Engineers Interested in AI/ML
- Researchers & Business Analysts

Tools and Lab Platform

- Python (Pandas, NumPy, Matplotlib, Seaborn, Scikit-learn, Statsmodels)
- Jupyter Notebook/Colab

Course Agenda

Module 1: Foundations of Statistical Analysis

- Role of Statistics in AI/ML
- Types of Data (Nominal, Ordinal, Interval, Ratio)
- Central Tendency (Mean, Median, Mode)
- Dispersion (Range, Variance, Std. Dev.)
- Data Visualization (Histogram, Boxplot, Scatterplot)
- Introduction to Probability

Practice exercises

- Descriptive statistics in Pandas
- Visualizations using Seaborn

- Exploratory Data Analysis (EDA) on real dataset

Module 2: Inferential Statistics & Hypothesis Testing

- Probability Distributions (Normal, Binomial, Poisson)
- Sampling Techniques
- Central Limit Theorem
- Confidence Intervals & Margin of Error
- Hypothesis Testing: t-test, ANOVA, Chi-square
- P-values, Significance Level, Type I/II Errors

Practice exercises

- Simulating distributions in Python
- Hypothesis testing using SciPy
- AB Testing Scenario (Email Campaign)

Module 3: Advanced Statistical Analysis for AI/ML

- Linear & Logistic Regression
- Correlation vs Causation
- Bayesian Statistics & Inference
- Dimensionality Reduction (PCA, t-SNE)
- Feature Engineering & Selection Techniques
- Cross-Validation & Grid Search CV

Practice exercises

- Regression with scikit-learn
- Feature selection via correlation, mutual info
- PCA on multidimensional dataset
- Cross-validation with GridSearch

Module 4: Time Series Analysis

- Components of Time Series (Trend, Seasonality, Noise)
- Stationarity & Differencing

- Autocorrelation (ACF, PACF)
- AR, MA, ARMA, ARIMA, SARIMA models
- Forecasting & Model Selection

Practice exercises

- Time Series Decomposition
- ACF/PACF plots with statsmodels
- Build & evaluate ARIMA model (Stock/Weather dataset)
- Project: Forecast electricity demand

Module 5: Real-World Applications & Model Evaluation

- Markov Chains & Hidden Markov Models
- Experimental Design & AB Testing
- Evaluation Metrics (Confusion Matrix, ROC, Precision-Recall, R^2, MSE)
- Business case mapping: Healthcare, Retail, Finance

Practice exercises

- Simulating Markov Chains in Python
- Confusion Matrix & ROC Curve
- Mini Project: Customer churn prediction
- Capstone Project with data storytelling

Case Studies

- **Customer Demographics Analysis in Retail** – Segment customers based on purchasing patterns using summary stats & charts
- **A/B Testing for Ad Campaign Effectiveness** – Validate if the new campaign performs better using statistical tests
- **Employee Attrition Prediction** – Use logistic regression and feature engineering to identify why employees leave
- **Forecasting Airline Passenger Demand** – Use ARIMA to forecast ticket sales based on historical data
- **Customer Churn Prediction for Telecom** – Build and evaluate a model to reduce churn using statistical evaluation metrics.