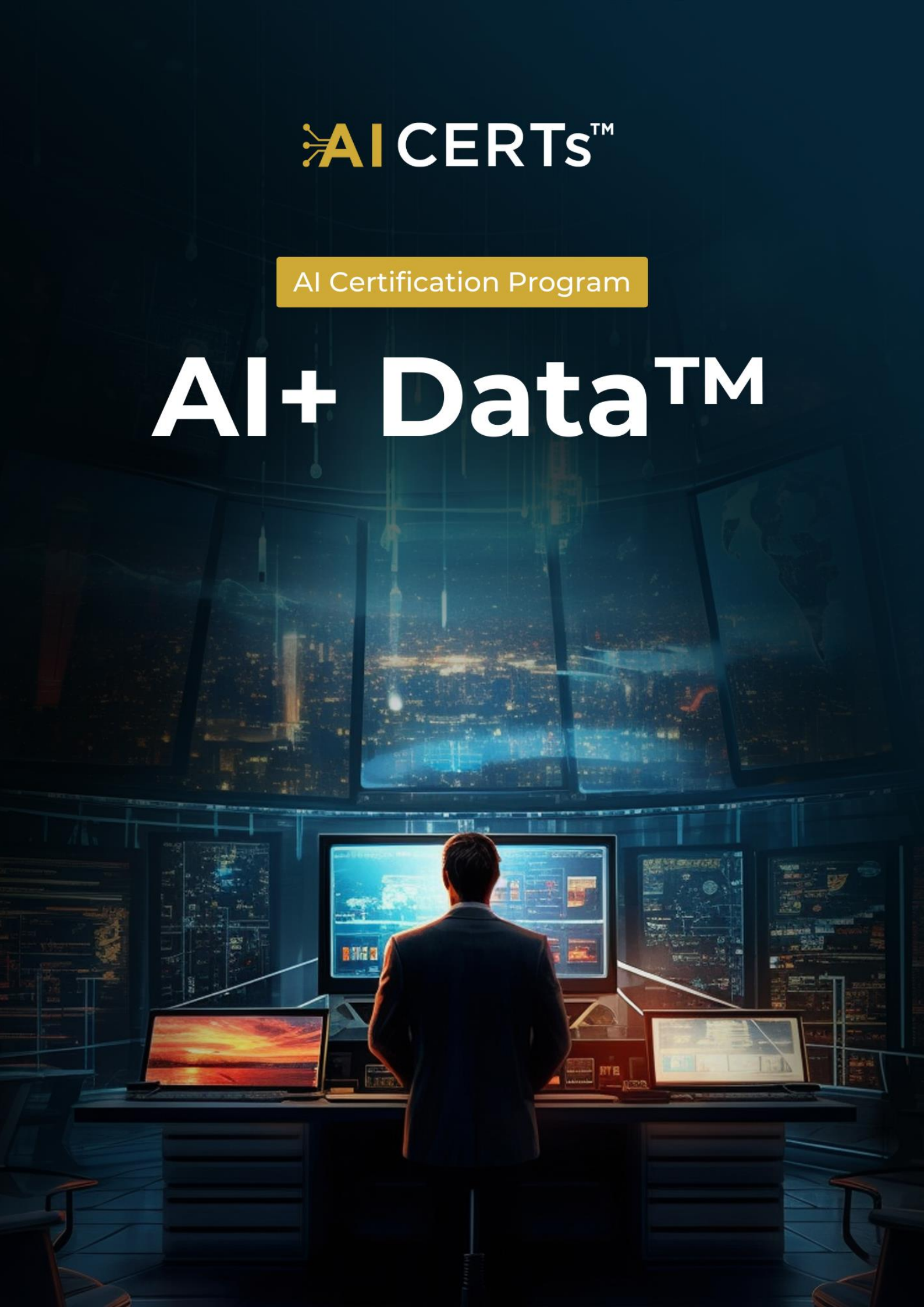
AI CERTs™

AI Certification Program

AI+ Data™



AI+ Data Certification - 40-Hour Training Plan

Total Duration: 40 hours

Format: Interactive Workshop (Presentations, Discussions, Case Studies)

Certification Modules

Module 1: Foundations of Data Science

- Introduction to Data Science
- Data Science Life Cycle
- Applications of Data Science

Module 2: Foundations of Statistics

- Basic Concepts of Statistics
- Probability Theory
- Statistical Inference

Module 3: Data Sources and Types

- Types of Data
- Data Sources
- Data Storage Technologies

Module 4: Programming Skills for Data Science

- Introduction to Python for Data Science
- Introduction to R for Data Science

Module 5: Data Wrangling and Preprocessing

- Data Imputation Techniques
- Handling Outliers and Data Transformation

Module 6: Exploratory Data Analysis (EDA)

- Introduction to EDA
- Data Visualization

Module 7: Generative AI Tools for Deriving Insights

- Introduction to Generative AI Tools
- Applications of Generative AI

Module 8: Machine Learning

- Introduction to Supervised Learning Algorithms

- Introduction to Unsupervised Learning
- Different Algorithms for Clustering
- Association Rule Learning with Implementation

Module 9: Advanced Machine Learning

- Ensemble Learning Techniques
- Dimensionality Reduction
- Advanced Optimization Techniques

Module 10 : Data-Driven Decision-Making

- Introduction to Data-Driven Decision Making
- Open-Source Tools for Data-Driven Decision Making
- Deriving Data-Driven Insights from Sales Dataset

Module 11 : Data Storytelling

- Understanding the Power of Data Storytelling
- Identifying Use Cases and Business Relevance
- Crafting Compelling Narratives
- Visualizing Data for Impact

Module 12 : Capstone Project - Employee Attrition Prediction

- Project Introduction and Problem Statement
- Data Collection and Preparation
- Data Analysis and Modeling
- Data Storytelling and Presentation

Module 1: Foundations of Data Science

This session planner offers a structured agenda for a 2-hour workshop on Module 1: Foundations of Data Science, covering key concepts, practical exercises, and discussions to deepen participants' understanding of data science principles and applications.

1. Introduction to Data Science (30 minutes)

- Define Data Science and its significance in various industries
- Discuss the role of data scientists and their responsibilities
- Explore the importance of data processing programming languages: Python, R, and SQL
- Highlight the systematic procedure of data manipulation and analysis

2. Data Science Life Cycle (30 minutes)

- Introduce the data science life cycle and its stages
- Explain the key steps involved in each stage of the data science process
- Emphasize the importance of problem definition, data preparation, modeling, evaluation, and deployment
- Discuss the iterative nature of the data science life cycle and the need for continuous improvement

3. Applications of Data Science (40 minutes)

- Explore real-world applications of data science in different industries such as finance, healthcare, e-commerce, and more
- Present case studies showcasing how data science is used for decision-making, predictive analytics, and problem-solving
- Discuss the impact of data science on business operations, customer insights, and strategic planning
- Highlight the transformative power of data science in addressing global challenges and driving innovation

4. Hands-on Exercise: Data Analysis Project (20 minutes)

- Engage participants in a hands-on data analysis project using a sample dataset
- Guide participants through the data science life cycle stages, from problem definition to model evaluation
- Encourage participants to apply their knowledge of data science concepts to solve a practical data analysis challenge

5. Data Science Ethics Discussion (20 minutes)

- Address the ethical considerations in data science and the importance of responsible data practices
- Discuss ethical dilemmas that data scientists may encounter in their work
- Encourage participants to reflect on the ethical implications of data collection, analysis, and decision-making

6. Q&A and Wrap-Up (20 minutes)

- Allow participants to ask questions and share their insights from the session
- Summarize key takeaways from the session on the foundations of data science
- Provide resources for further learning and exploration in the field of data science

Module 2: Foundations of Statistics

This session planner outlines a structured agenda for a 2-hour workshop on Module 2: Foundations of Statistics, covering key statistical concepts, probability theory, and statistical inference through a mix of theory, practical exercises, and discussions to enhance participants' understanding of statistical principles and applications.

1. Basic Concepts of Statistics (40 minutes)

- Overview of Descriptive Statistics: Importance and Significance
- Explanation of Descriptive Statistics Measures and their practical applications
- Discussion on Central Tendency Measures: Mean, Median, Mode
- Calculations and Interpretations for Variability Measures: Range, Variance, Standard Deviation, Interquartile Range (IQR)
- Introduction to Shape and Frequency Distribution Measures: Skewness, Kurtosis, Percentiles, Frequency Distributions
- Real-world examples and case studies to illustrate statistical concepts

2. Probability Theory (30 minutes)

- Introduction to Probability Theory and its relevance in statistics
- Explanation of basic probability concepts: events, sample spaces, and probabilities
- Discussion on probability distributions: discrete and continuous distributions
- Calculating probabilities and understanding the rules of probability
- Application of probability theory in data analysis and decision-making

3. Statistical Inference (40 minutes)

- Overview of Statistical Inference and its importance in drawing conclusions from data
- Introduction to Hypothesis Testing: Null and Alternative Hypotheses
- Explanation of Confidence Intervals and their interpretation
- Conducting hypothesis tests using statistical tests such as t-tests and chi-square tests
- Practich

5. Discussion and Application (20 minutes)

- Facilitate a discussion on the practical applications of statistics in various fields
- Encourage participants to share their experiences with applying statistical concepts in their work or studies

- Discuss the role of statistics in decision-making, research, and problem-solving

6. Q&A and Wrap-Up (10 minutes)

- Allow participants to ask questions and seek clarification on statistical concepts covered
- Summarize key takeaways from the session on the foundations of statistics
- Provide resources for further learning and practice in statistics

Module 3: Data Sources and Types

This session planner provides a comprehensive agenda for a 4-hour workshop on Module 3: Data Sources and Types, covering different data types, sources, and storage technologies through a mix of theory, hands-on activities, and group discussions to deepen participants' understanding of data management concepts.

1. Types of Data (1 hour)

- Introduction to Structured, Semi-structured, and Unstructured Data
- Explanation of the characteristics and features of each data type
- Discussion on the importance of data classification for analysis and storage
- Examples of structured data schemas and their applications
- Case studies showcasing the use of semi-structured and unstructured data in different industries

2. Data Sources (1 hour)

- Overview of Data Sources and their significance in data collection
- Discussion on Internal and External Data Sources
- Exploration of primary and secondary data sources
- Consideration of data quality issues and data governance practices
- Hands-on activity: Identifying data sources for a given scenario

3. Data Storage Technologies (1 hour)

- Introduction to Data Storage Technologies: Databases, Data Lakes, and Data Warehouses
- Explanation of Relational Databases and NoSQL databases
- Comparison of SQL and NoSQL databases in terms of scalability and flexibility
- Overview of Data Lake architecture and its advantages for storing large volumes of data
- Practical demonstration of querying data from different storage technologies

4. Data Management and Analysis Workshop (1 hour)

- Hands-on Data Management Exercise: Cleaning and Preparing Data for Analysis
- Data Visualization Exercise: Creating visualizations from different data types
- Data Analysis Exercise: Applying statistical techniques to analyze data from various sources
- Group Discussion: Sharing insights and challenges faced during data management and analysis

5. Breaks and Networking Opportunities (Throughout)

- Scheduled breaks for participants to refresh and network with peers
- Informal networking sessions to encourage collaboration and knowledge sharing

6. Q&A and Wrap-Up (30 minutes)

- Open floor for participants to ask questions and seek clarification on data sources and types
- Summary of key learnings from the session on data sources and storage technologies
- Distribution of resources for further exploration and learning in the field of data management

Module 4: Programming Skills for Data Science

This session planner outlines a detailed agenda for a 6-hour workshop on Module 4: Programming Skills for Data Science, covering introductory and advanced concepts in Python and R programming for data analysis and visualization, with a focus on hands-on exercises and practical applications to enhance participants' programming skills in data science.

1. Introduction to Python for Data Science (2 hours)

- Overview of Python programming language and its role in data science
- Introduction to Python syntax, data types, and variables
- Hands-on Python coding exercises for beginners
- Exploring popular Python data analysis libraries: Pandas, NumPy, SciPy
- Practical examples of data manipulation and visualization using Python

2. Intermediate Python Concepts and Libraries (1.5 hours)

- Advanced Python programming concepts: Functions, Modules, and Packages
- Introduction to data visualization with Matplotlib and Seaborn
- Implementing machine learning algorithms with Scikit-learn
- Hands-on exercises to apply Python libraries for data analysis and visualization

3. Introduction to R for Data Science (2 hours)

- Overview of the R programming language and its applications in data science
- Introduction to R syntax, data structures, and functions
- Hands-on R coding exercises for data manipulation and analysis
- Exploring popular R packages for statistical analysis and visualization
- Practical examples of data wrangling and exploratory data analysis in R

4. Advanced R Programming and Data Visualization (1.5 hours)

- Advanced R programming techniques: Control Structures, Functions, and Loops
- Introduction to advanced data visualization in R using ggplot2
- Applying statistical techniques in R for hypothesis testing and regression analysis
- Hands-on exercises to practice advanced R programming and data visualization

5. Breaks and Networking Opportunities (Throughout)

- Scheduled breaks for participants to rest and network with peers
- Informal networking sessions to encourage collaboration and knowledge sharing

6. Q&A and Wrap-Up (30 minutes)

- Open floor for participants to ask questions and seek clarification on Python and R programming concepts
- Summary of key takeaways from the session on programming skills for data science
- Distribution of resources for further learning and practice in Python and R for data science

Module 5: Data Wrangling and Preprocessing

This session planner provides a structured agenda for a 4-hour workshop on Module 5: Data Wrangling and Preprocessing, focusing on data imputation techniques, outlier handling, data transformation, and advanced data preprocessing concepts using Python libraries. The hands-on exercises and group activities aim to enhance participants' skills in preparing and cleaning data for analysis and machine learning applications.

1. Data Imputation Techniques (2 hours)

- Introduction to Missing Data and its impact on data analysis
- Overview of Data Imputation methods: Mean, Median, Mode, Regression, KNN
- Comparison of different imputation techniques and their suitability for various scenarios
- Hands-on exercises to practice data imputation using Python libraries like Pandas and Scikit-learn
- Discussion on the importance of choosing the right imputation technique based on data characteristics

2. Handling Outliers and Data Transformation (1.5 hours)

- Understanding Outliers and their effects on statistical analysis
- Techniques for detecting and handling outliers: Z-score, IQR, Boxplot
- Data Transformation methods: Normalization and Standardization
- Practical examples of outlier detection and data transformation using Python libraries
- Group activity: Identifying outliers in a dataset and applying data transformation techniques

3. Advanced Data Preprocessing Techniques (0.5 hours)

- Introduction to Feature Engineering and its role in improving model performance
- Overview of Encoding Categorical Variables and its importance in machine learning
- Hands-on demonstration of encoding categorical variables using Python libraries
- Discussion on the challenges and best practices in data preprocessing for machine learning models

4. Breaks and Networking Opportunities (Throughout)

- Scheduled breaks for participants to relax and network with peers
- Informal networking sessions to encourage collaboration and knowledge sharing

5. Q&A and Wrap-Up (30 minutes)

- Open discussion for participants to ask questions and seek clarification on data wrangling and preprocessing techniques

- Summary of key learnings from the session on data imputation, outlier handling, and data transformation
- Distribution of resources for further exploration and practice in data preprocessing techniques

Module 6: Exploratory Data Analysis (EDA)

This session planner outlines a comprehensive agenda for a 4-hour workshop on Module 6: Exploratory Data Analysis, covering introductory and advanced concepts in EDA and data visualization. The hands-on exercises, group activities, and discussions aim to enhance participants' skills in exploring and visualizing data to derive meaningful insights for decision-making and model building.

1. Introduction to EDA (2 hours)

- Overview of EDA and its significance in the data analysis process
- Understanding the goals and objectives of EDA in uncovering insights from data
- Exploring the role of EDA in identifying patterns, trends, and relationships in datasets
- Hands-on exercises to practice basic EDA techniques using Python and R
- Discussion on the importance of EDA in informing decision-making and model building

2. Data Visualization Techniques (1.5 hours)

- Introduction to Data Visualization and its importance in communicating insights
- Overview of popular data visualization libraries: Matplotlib, Seaborn, ggplot2
- Creating various types of visualizations: Line plots, Bar charts, Scatter plots
- Hands-on session to create visualizations and interpret data patterns
- Group activity: Visualizing a dataset and presenting findings to the group

3. Advanced Data Visualization and Interpretation (0.5 hours)

- Exploring advanced data visualization techniques: Heatmaps, Box plots, Pair plots
- Understanding the use of visualizations for feature selection and model evaluation
- Practical examples of using advanced visualizations to gain deeper insights from data
- Discussion on best practices for effective data visualization in EDA

4. Breaks and Networking Opportunities (Throughout)

- Scheduled breaks for participants to relax and network with peers
- Informal networking sessions to encourage collaboration and knowledge sharing

5. Q&A and Wrap-Up (30 minutes)

- Open forum for participants to ask questions and seek clarification on EDA and data visualization techniques
- Summary of key takeaways from the session on EDA and data visualization

- Distribution of resources for further learning and practice in exploratory data analysis and visualization

Module 7: Generative AI Tools for Deriving Insights

This session plan aims to provide participants with a comprehensive understanding of generative AI tools, their applications, and hands-on experience to derive insights from data using these advanced techniques.

1. Introduction to Generative AI Tools (1 hour)

- Overview of Generative AI and its significance in data science
- Explanation of key generative AI techniques: autoencoders, GANs, VAEs, transformers, Markov models, and deep belief networks (DBNs) [T6]
- Understanding how these techniques enable machines to learn and create new data instances across various domains and applications

2. Hands-on Practice with Generative AI Tools (1 hour)

- Practical applications of generative AI techniques such as autoencoders, GANs, VAEs, and transformers
- Using Python scripts and demos to apply these methods to real-world datasets and tasks
- Generating insights from data through generative AI tools for data analysis, synthesis, and visualization [T4]

3. Applications of Generative AI (1 hour)

- Exploring real-world applications of generative AI in various industries
- Discussing data generation, synthesis, and augmentation using generative models
- Understanding how generative AI can be used for image-to-image translation, text generation, NLP, anomaly detection, data denoising, and recommendation systems

Module 8: Machine Learning Refresher

This module encourage participants to explore further topics in machine learning and apply the learned concepts in their projects or work environents

1. Introduction to Supervised Learning Algorithms (1 hour)

- Definition and relevance of supervised learning in machine learning
- Explanation of simple linear regression as a basic supervised learning algorithm [T5]
- Understanding predictor and response variables in supervised learning
- Overview of the simple linear regression equation and its components

2. Applications to Unsupervised Learning (1 hour)

- Introduction to unsupervised learning and its applications in data analysis
- Explanation of clustering as a common unsupervised learning technique
- Discussing the differences between supervised and unsupervised learning approaches
- Practical examples of unsupervised learning applications in real-world scenarios

3. Different Algorithms for Clustering (1 hour)

- Overview of various clustering algorithms such as K-means, hierarchical clustering, and DBSCAN
- Understanding the principles and use cases of different clustering algorithms
- Hands-on practice with clustering algorithms using sample datasets
- Discussion on the importance of clustering in data segmentation and pattern recognition

4. Association Rule Learning (1 hour)

- Introduction to association rule learning and its significance in data mining
- Explanation of Apriori algorithm for association rule learning
- Practical examples of association rule learning in market basket analysis and recommendation systems
- Hands-on exercise to implement association rule learning using a sample dataset

Conclusion:

Recap of key concepts covered in the session

Q&A session to clarify doubts and discuss practical applications of machine learning algorithms

Module 9: Advanced Machine Learning

This module allow participants to explore further applications of ensemble learning, dimensionality reduction, and optimization techniques in their projects or research endeavors

1. Ensemble Learning Techniques (1 hour)

- Introduction to ensemble learning and its significance in improving predictive performance
- Overview of bagging, boosting, and stacking ensemble algorithms
- Explanation of bagging technique, including bootstrap sampling and combining predictions
- Discussion on boosting algorithms such as AdaBoost, Gradient Boosting Machines (GBM), XGBoost, and LightGBM
- Hands-on practice with ensemble learning techniques using sample datasets

2. Dimensionality Reduction (1 hour)

- Understanding the importance of dimensionality reduction in machine learning
- Explanation of techniques like Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE)
- Practical examples of dimensionality reduction for feature selection and visualization
- Hands-on exercise to implement dimensionality reduction on a dataset and interpret the results

3. Advanced Optimization Techniques (1 hour)

- Introduction to advanced optimization techniques in machine learning
- Explanation of optimization algorithms like stochastic gradient descent, Adam, RMSprop, and more
- Understanding the role of optimization in training deep learning models
- Hands-on practice with optimizing machine learning models using different algorithms

Conclusion

Recap of key concepts covered in the session

Q&A session to address any queries or concerns related to advanced machine learning techniques

Module 10: Data-Driven Decision-Making

This module encourage participants to explore importance of data-driven decision-making and its role in various business sectors and its operations.

1. Introduction to Data-Driven Decision Making (1 hour)

- Definition and importance of data-driven decision-making in modern business strategies
- Discussing the concepts, benefits, and applications of data-driven decision-making
- Understanding the role of data in improving outcomes, strategic planning, and innovation
- Real-world examples of successful data-driven decision-making processes

2. Open-Source Tools for Data-Driven Decision Making (1 hour)

- Overview of popular open-source tools for data analysis and decision-making such as R, Python, and Tableau
- Hands-on demonstration of using open-source tools for data visualization, analysis, and interpretation
- Exploring the capabilities of tools like Jupyter Notebooks, Pandas, and Matplotlib for data-driven insights
- Practical exercises to apply open-source tools to real-world datasets

3. Deriving Data-Driven Insights from Sales Dataset (1 hour)

- Case study on deriving insights from a sales dataset using data-driven approaches
- Data exploration, analysis, and visualization techniques for extracting meaningful insights
- Identifying patterns, trends, and correlations in the sales data to make informed decisions
- Group discussion on the key insights derived from the sales dataset and their implications

Module 11: Data Storytelling

This session plan is designed to provide participants with a focused and interactive learning experience on data storytelling, covering key aspects such as understanding the power of storytelling, identifying use cases, crafting narratives, and visualizing data for impact within a 2-hour timeframe.

1. Understanding the Power of Data Storytelling (30 minutes)

- Introduction to the concept of data storytelling and its significance in conveying insights effectively
- Exploring how data storytelling can transform complex data into compelling narratives
- Discussing the impact of storytelling on decision-making processes and stakeholder engagement
- Examples of successful data storytelling initiatives in various industries

2. Identifying Use Cases and Business Relevance (30 minutes)

- Discussing the importance of identifying relevant use cases for data storytelling in business contexts
- Brainstorming potential scenarios where data storytelling can drive business value and decision-making
- Identifying key data points and metrics that align with business objectives for effective storytelling
- Group activity to analyze and prioritize use cases based on their impact and relevance

3. Crafting Compelling Narratives (30 minutes)

- Techniques for crafting engaging and impactful narratives around data insights
- Understanding the elements of a compelling data story, including structure, characters, and conflict
- Tips for aligning data narratives with audience preferences and communication goals
- Hands-on exercise to draft a data story outline based on a given dataset or scenario

4. Visualizing Data for Impact (30 minutes)

- Importance of data visualization in enhancing the impact of data storytelling
- Best practices for creating visually appealing and informative data visualizations
- Overview of visualization tools and techniques for conveying data insights effectively
- Practical demonstration of creating data visualizations using tools like Tableau or Power BI

Conclusion

- Recap of key concepts covered in the session
- Q&A session to address any queries or concerns related to data storytelling
- Encouraging participants to apply data storytelling techniques in their projects or presentations

Module 12: Capstone Project - Employee Attrition Prediction

This session plan aims to provide participants with a comprehensive understanding of the capstone project on employee attrition prediction, including the project introduction, problem statement, data collection, and preparation within a 2-hour timeframe.

1. Project Introduction and Problem Statement (1 hour)

- Overview of the capstone project on employee attrition prediction
- Introduction to the problem statement and objectives of the project
- Discussing the significance of predicting employee attrition for organizational success
- Presenting real-world examples of the impact of high attrition rates on businesses

2. Data Collection and Preparation (1 hour)

- Explanation of the data collection process for the employee attrition prediction project
- Identifying relevant data sources and variables for predicting attrition
- Hands-on session on data cleaning, preprocessing, and feature engineering techniques
- Group activity to explore and analyze the dataset for potential insights and patterns

Conclusion

- Recap of key points covered in the session
- Setting expectations for the next steps in the capstone project
- Encouraging participants to start working on data collection and preparation tasks

Contact

252 West 37th St., Suite 1200W
New York, NY 10018