



AiNS Computer Institute

DATA SCIENCE

"Discover. Analyze. Transform: Data
Science Essentials."

JOIN
AiNS



+91 8805845720

ainscomputerinstitute.com



DATA SCIENCE?

Data Science is an interdisciplinary field that utilizes scientific methods, algorithms, processes, and systems to extract insights and knowledge from structured and unstructured data. It combines aspects of statistics, mathematics, computer science, and domain expertise to understand complex phenomena, make predictions, and drive decision-making.

Key components of data science include data collection, data cleaning and preprocessing, exploratory data analysis, machine learning and statistical modeling, data visualization, and interpretation of results. Data scientists often work with large datasets, using programming languages like Python, SQL, along with specialized tools and frameworks for data manipulation and analysis.

Applications of data science span across various industries such as healthcare, finance, marketing, retail, and more. It plays a crucial role in areas like predictive analytics, recommendation systems, fraud detection, natural language processing, and image recognition.

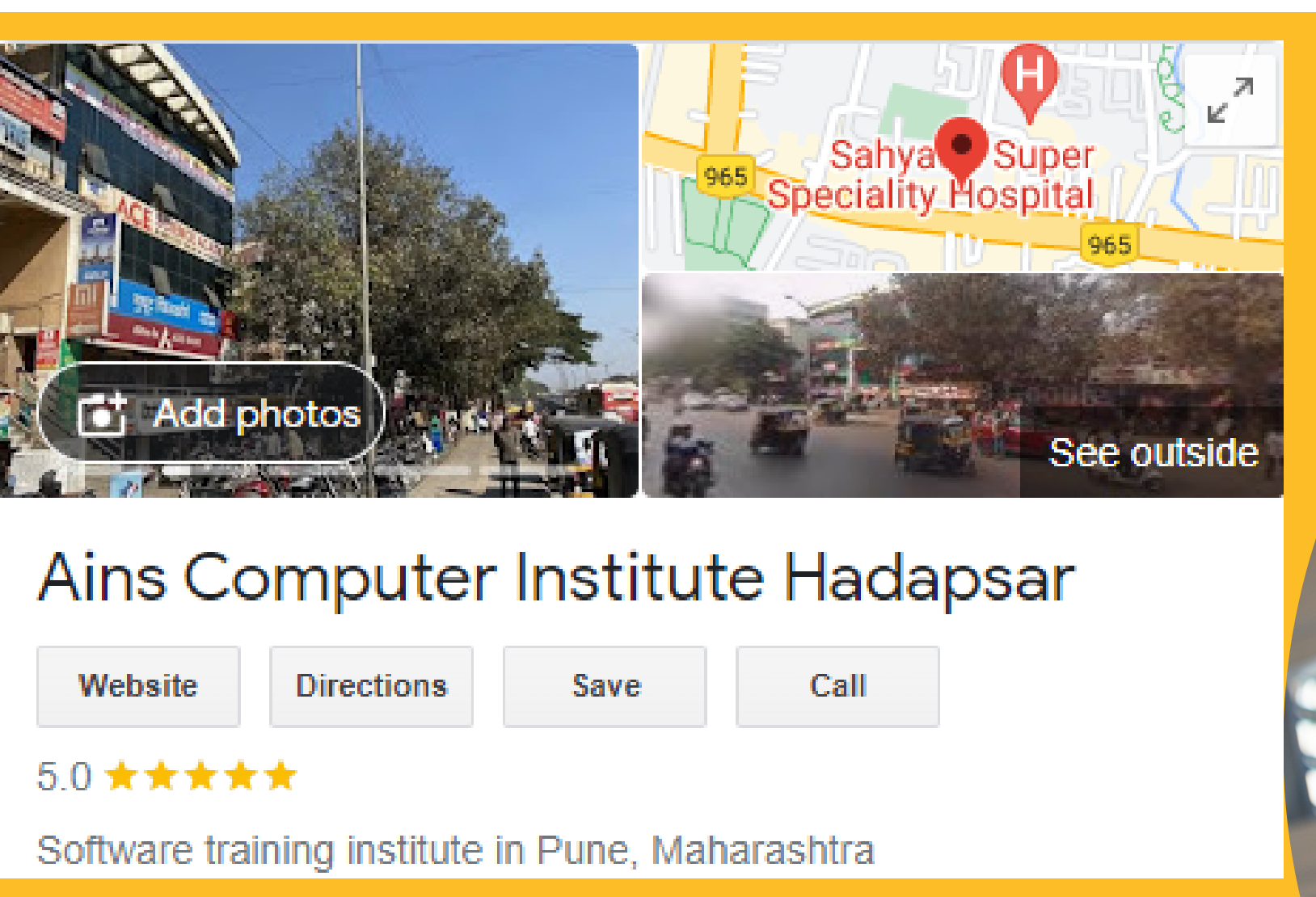
Overall, data science aims to extract actionable insights and create value from data to solve complex problems and drive innovation in organizations.



WHY AiNS?

At AiNS Computer Institute, our dedicated team of faculty members who teach Data Science are the backbone of our program. They bring a wealth of experience, expertise, and passion to the classroom, inspiring and guiding our students on their journey to mastering this complex and rapidly evolving field.

Our Data Science faculty members are highly skilled professionals with extensive backgrounds in statistics, mathematics, computer science, and related disciplines. They possess a deep understanding of the latest tools, techniques, and methodologies used in the industry, ensuring that our curriculum remains relevant and up-to-date.



A screenshot of a Google Maps listing for 'Ains Computer Institute Hadapsar'. The listing includes a street view image on the left, a map snippet in the top right showing the location near 'Sahya Super Speciality Hospital', and a street view image in the bottom right. Below the images, the name 'Ains Computer Institute Hadapsar' is displayed, followed by buttons for 'Website', 'Directions', 'Save', and 'Call'. The rating is 5.0 stars, and the description is 'Software training institute in Pune, Maharashtra'.

Ains Computer Institute Hadapsar

Website Directions Save Call

5.0 ★★★★★

Software training institute in Pune, Maharashtra



BENEFITS at AiNS



1

Expert Faculty

2

Industry-Relevant Curriculum

3

Career Guidance & Opportunities

4

Recognized ISO Certification

5

Real-World Data Analysis Projects

6

Access to Premium Tools and Datasets

7

Get Interview Ready & Resume Building

8

100% Job Assistance



Certificate of Completion

This Certificate is presented to SAMPLE for successfully completing

his/her course in SAMPLE by securing 99%

We wish him/her success in life.

*(This percentage is based on student's Practical
& Written examination)*

Digitally signed by

Nasir Hassan Nasir Hassan Khan
Khan
Date: 2024.03.16
16:08:40 +05'30'

AUTHORISED SIGNATORY
DIRECTOR



9001:2015



OUR CERTIFICATION:

ISO course certification offers several advantages:

1. **International Recognition:** ISO certifications are globally recognized, demonstrating compliance with international standards and best practices.
2. **Enhanced Credibility:** Certification enhances the credibility and reputation of the organization, instilling confidence in customers, stakeholders, and partners.
3. **Improved Efficiency:** ISO standards often focus on improving processes, leading to increased efficiency, reduced waste, and cost savings.
4. **Market Access:** Certification can provide easier access to new markets, as many contracts and tenders require ISO certification as a prerequisite.
5. **Risk Management:** ISO standards help identify and mitigate risks, ensuring better management of potential threats to the organization.
6. **Customer Satisfaction:** Compliance with ISO standards often leads to higher customer satisfaction due to improved quality and consistency of products or services.
7. **Competitive Advantage:** Certification can give a competitive edge by demonstrating commitment to quality, safety, and environmental responsibility.
8. **Legal Compliance:** ISO standards often align with legal requirements, ensuring compliance and reducing the risk of penalties or legal issues.
9. **Continuous Improvement:** ISO certification encourages a culture of continuous improvement, leading to ongoing refinement of processes and performance.
10. **Employee Morale:** Employees often take pride in working for a certified organization, leading to increased morale and motivation.

value

adding

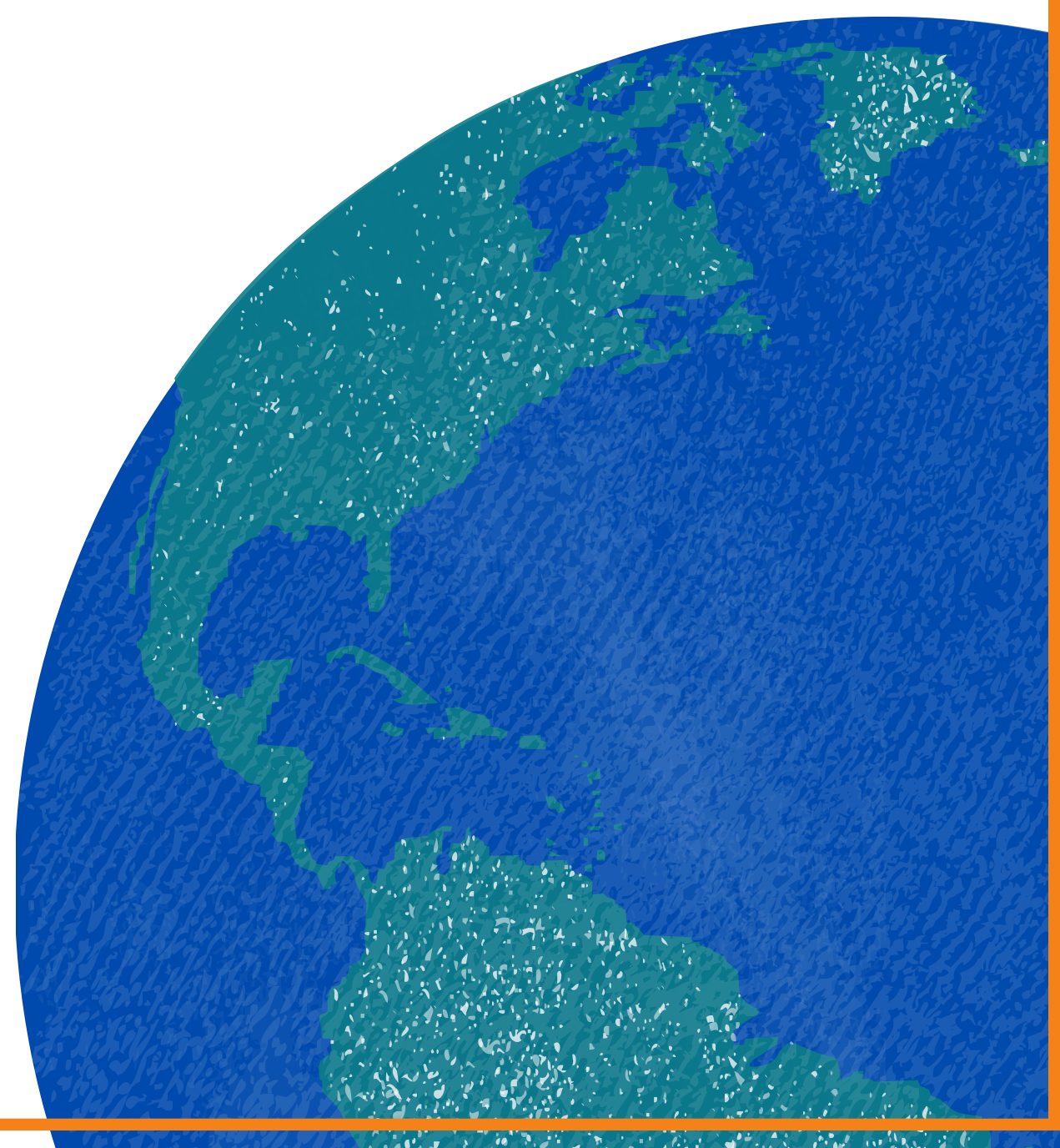
DEMAND:

In India:

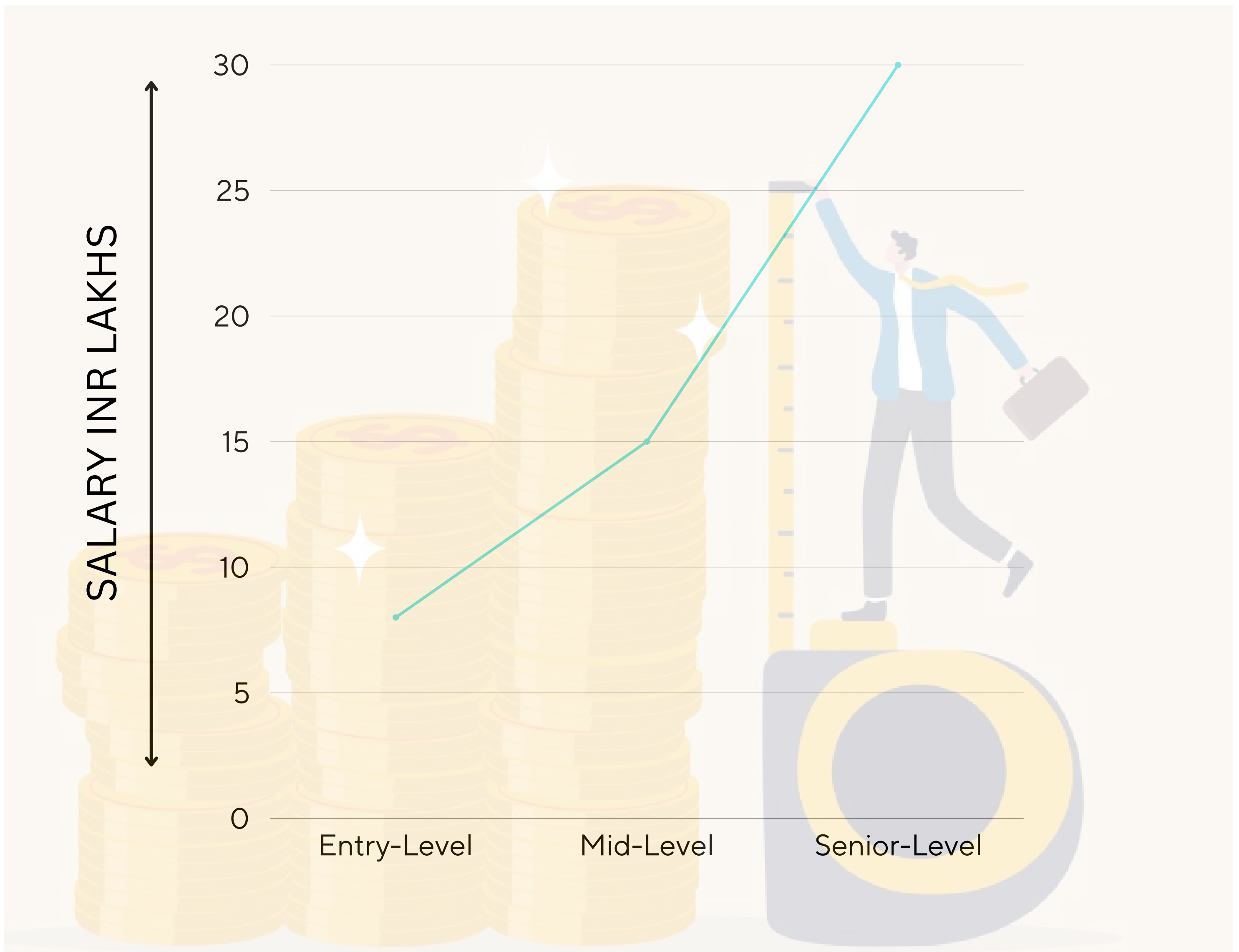
1. The demand for Data Science professionals has been growing rapidly, with reports indicating a year-on-year increase in demand.
2. Various studies and industry reports have estimated the annual growth rate of demand for Data Science professionals in India to be over 30%.
3. The demand for Data Science talent is particularly high in cities such as Bangalore, Hyderabad, Mumbai, Pune, and Delhi NCR, which are major hubs for technology and analytics companies.

Globally:

1. The demand for Data Science professionals is also high on a global scale, with companies across industries recognizing the importance of data-driven decision-making.
2. Data Science-related roles consistently rank among the top in-demand jobs globally, according to reports from sources such as LinkedIn, Glassdoor, and Indeed.
3. The demand for Data Science talent is particularly strong in technology hubs such as Silicon Valley in the United States, London in the United Kingdom, and cities in countries like Canada, Australia, and Singapore.



SALARY IN INDIA:



Global:

- In the **United States**, the average annual salary for Data Scientists ranges from **\$80,000 to \$160,000**, depending on factors such as location (e.g., Silicon Valley tends to have higher salaries), industry, and level of experience.
- In the **United Kingdom**, Data Scientists earn average annual salaries ranging from **£30,000 to £70,000**, with higher salaries in cities like London.
- In **Canada**, the average annual salary for Data Scientists ranges from **CAD 60,000 to CAD 120,000**, depending on factors such as location, industry, and experience.
- In **Australia**, Data Scientists earn average annual salaries ranging from **AUD 80,000 to AUD 150,000**, with higher salaries in cities like Sydney and Melbourne.
- In **Singapore**, Data Scientists earn average annual salaries ranging from **SGD 50,000 to SGD 150,000**, depending on factors such as experience, industry, and company size.

TOP COMPANIES HIRING DATA SCIENCE:



& MANY MORE



Curriculum

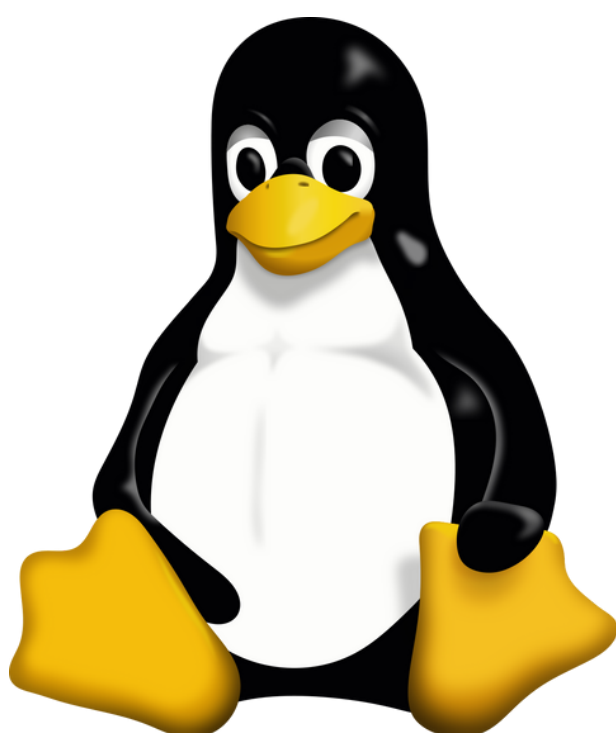


Module 1 – Preparatory Session - Linux and Python Python

- Introduction to Python and IDEs – The basics of the Python programming language, how you can use various IDEs for python development like Jupyter, Pycharm, etc.
- Python Basics – Variables, Data Types, Loops, Conditional Statements, functions, decorators, lambda functions, file handling, exception handling ,etc.
- Object Oriented Programming – Introduction to OOPs concepts like classes, objects, inheritance, abstraction, polymorphism, encapsulation, etc.
- Hands-on Sessions And Assignments for Practice – The culmination of all the above concepts with real-world problem statements for better understanding.

Linux

- Introduction to Linux – Establishing the fundamental knowledge of how Linux works and how you can begin with Linux OS.
- Linux Basics – File Handling, data extraction, etc.
- Hands-on Sessions And Assignments for Practice – Strategically curated problem statements for you to start with Linux.



Module 2 – Data Wrangling with SQL

SQL Basics –

- Fundamentals of Structured Query Language
- SQL Tables, Joins, Variables

Advanced SQL –

- SQL Functions, Subqueries, Rules, Views
- Nested Queries, string functions, pattern matching
- Mathematical functions, Date-time functions, etc.

Deep Dive into User Defined Functions

- Types of UDFs, Inline table value, multi-statement table.
- Stored procedures, rank function, SQL ROLLUP, etc.

SQL Optimization and Performance

- Record grouping, searching, sorting, etc.
- Clustered indexes, common table expressions.

Hands-on exercise:

Writing comparison data between the past year and the present year with respect to top products, ignoring the redundant/junk data, identifying the meaningful data, and identifying the demand in the future(using complex subqueries, functions, pattern matching concepts).



Module 3 – Python With Data Science

Extract Transform Load



● Interacting with APIs

Data Handling with NumPy

- NumPy Arrays, CRUD Operations, etc.
- Linear Algebra – Matrix multiplication, CRUD operations, Inverse, Transpose, Rank, Determinant of a matrix, Scalars, Vectors, Matrices.

Data Manipulation Using Pandas

- Loading the data, data frames, series, CRUD operations, splitting the data, etc.

Data Preprocessing

- Exploratory Data Analysis, Feature engineering, Feature scaling, Normalization, standardization, etc.
- Null Value Imputations, Outliers Analysis and Handling, VIF, Bias-variance trade-off, cross-validation techniques, train-test split, etc.

Data Visualization

- Bar charts, scatter plots, count plots, line plots, pie charts, donut charts, etc. with Python matplotlib.
- Regression plots, categorical plots, area plots, etc, with Python seaborn.



Module 4 – Inferential Analytics

Statistics and Descriptive Analytics using MS Excel

- **Measure of central tendency, measure of spread, five points summary, etc.**
- **Probability Distributions, Probability in Business Analytics**
- **Probability Distributions, Binomial distribution, Poisson distribution, bayes theorem, central limit theorem.**

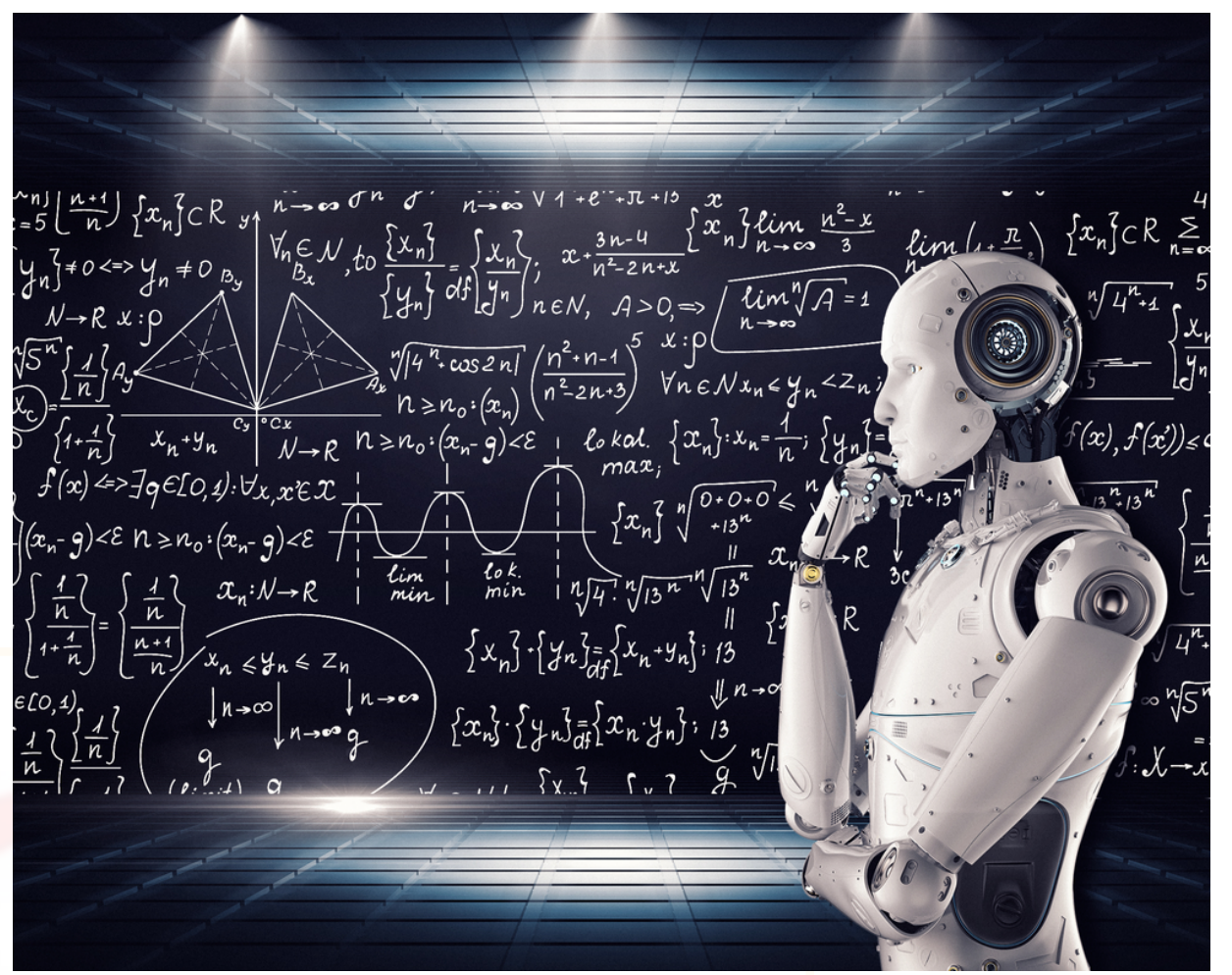
Python for Descriptive, Diagnostic, and Inferential Statistics

- **Correlation, covariance, confidence intervals, hypothesis testing, F-test, Z-test, t-test, ANOVA, chi-square test, etc.**

Module 5 – Machine Learning

Introduction to Machine Learning

- Supervised, Unsupervised Learning.
 - Introduction to scikit-learn, Keras, etc.
- #### Regression



- Introduction classification problems, Identification of a regression problem, dependent and independent variables.
- How to train the model in a regression problem.
- How to evaluate the model for a regression problem.
- How to optimize the efficiency of the regression model.

Classification

- Introduction to classification problems, Identification of a classification problem, and dependent and independent variables.
- How to train the model in a classification problem.
- How to evaluate the model for a classification problem.
- How to optimize the efficiency of the classification model.

Clustering

- Introduction to clustering problems, Identification of a clustering problem, dependent and independent variables.
- How to train the model in a clustering problem.
- How to evaluate the model for a clustering problem.
- How to optimize the efficiency of the clustering model.

Module 6 – Supervised Learning

- **Linear Regression** – Creating linear regression models for linear data using statistical tests, data preprocessing, standardization, normalization, etc.
- **Logistic Regression** – Creating logistic regression models for classification problems – such as if a person is diabetic or not, if there will be rain or not, etc.
- **Decision Tree** – Creating decision tree models on classification problems in a tree like format with optimal solutions.
- **Random Forest** – Creating random forest models for classification problems in a supervised learning approach.
- **Support Vector Machine** – SVM or support vector machines for regression and classification problems.
- **Gradient Descent** – Gradient descent algorithm that is an iterative optimization approach to finding the local minimum and maximum of a given function.
- **K-Nearest Neighbors** – A simple algorithm that can be used for classification problems.
- **Time Series Forecasting** – Making use of time series data, gathering insights and useful forecasting solutions using time series forecasting.



Module 7 – Unsupervised Learning

- **K-means** – The k-means algorithm that can be used for clustering problems in an unsupervised learning approach.
- **Dimensionality reduction** – Handling multi dimensional data and standardizing the features for easier computation.
- **Linear Discriminant Analysis** – LDA or linear discriminant analysis to reduce or optimize the dimensions in the multidimensional data.
- **Principal Component Analysis** – PCA follows the same approach in handling the multidimensional data.

Performance Metrics

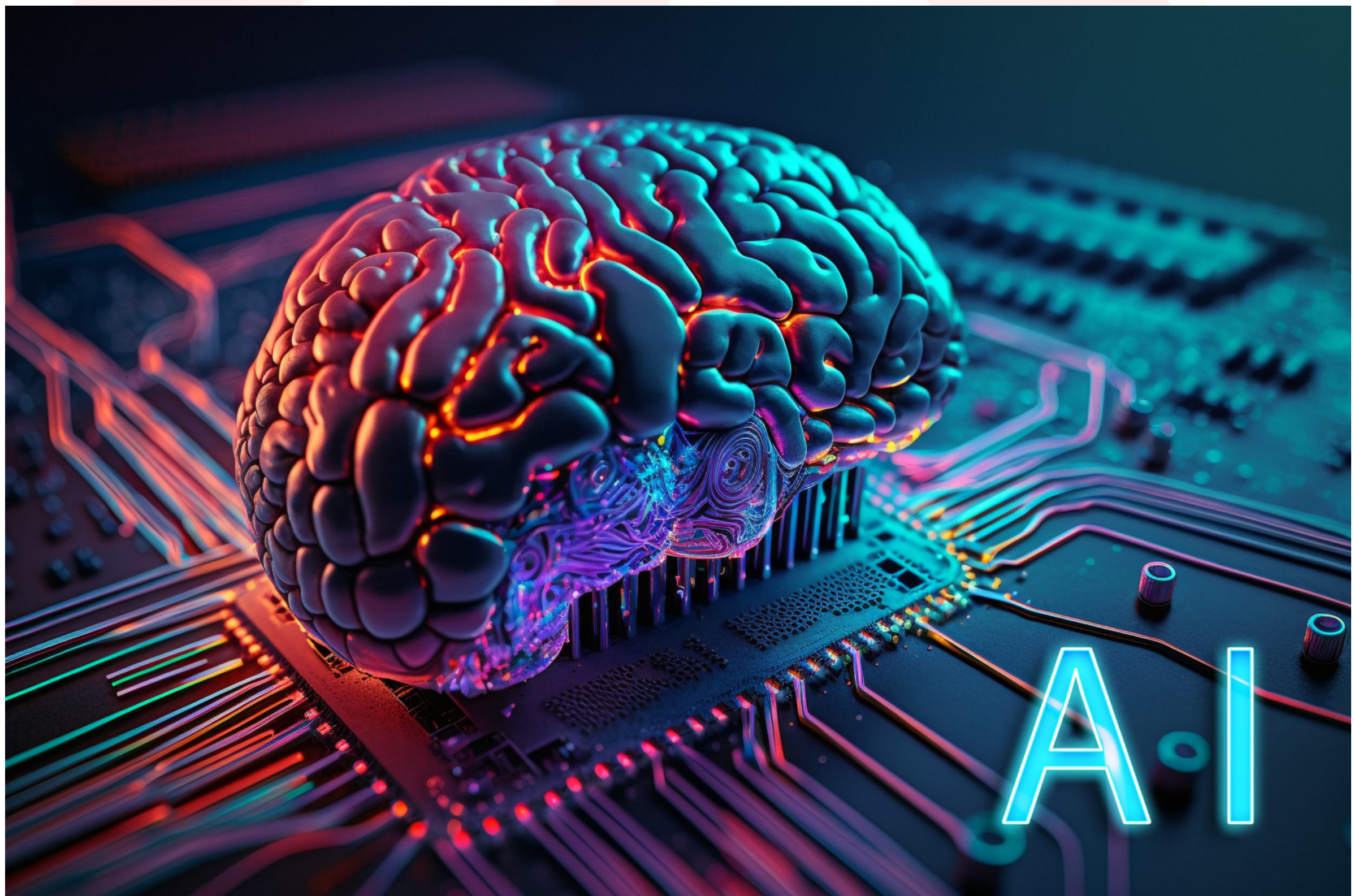
- **Classification reports** – To evaluate the model on various metrics like recall, precision, f-support, etc.
- **Confusion matrix** – To evaluate the true positive/negative, and false positive/negative outcomes in the model.
- **r2, adjusted r2, mean squared error, etc.**



Module 8 - Advanced Machine Learning Algorithms

Bagging And Boosting Algorithms

- **XGBoost** – Extreme Gradient Boosting or XGBoost algorithm works on an ensemble approach that combines the predictions of weak models to produce a strong prediction.
- **ADABOOST** – An ensemble machine learning algorithm that converts the weak learners to strong learners for better performance.
- **Gradient Tree Boosting** – One of the powerful boosting algorithms works on the principle of reducing the bias in predictions.
- **Bootstrap Aggregation** – A bagging technique that fits decision trees on bootstrap samples of the training data.



Other Machine Learning Algorithms

- **Ordinary Least Square** – A simple regression algorithm that is used to estimate the coefficient of the linear regression equation for predictions.
- **Markov Chain** – A stochastic approach or algorithm that describes the possible chain of events in terms of probability theory.
- **Naive Bayes** – Bayes theorem inspired algorithm in machine learning that is used for predictive modeling.
- **Stochastic Gradient Descent** – A variant of Gradient Descent that is used to optimize machine learning algorithms.
- **Gaussian Mixture Model** – A probabilistic model that assumes the data points are generated from a mix of gaussian distributions with unknown parameters.
- **Singular Value Decomposition** – Algorithm based on factorization of a real or complex matrix, that is often used to create a recommendation engine.

Predictive Analytics And Machine Learning

- **Regression and Multivariate Analysis**
- **Classification problems in machine learning**
- **Data Multidimensionality and Linear Algebra**
- **Feature engineering and Feature selection**
- **Hyperparameter Tuning and other optimization techniques**

Cognitive Science and Analytics

- **Understanding Natural Language Processing Applications (e.g. Search Engines and Social Media)**
- **Web Analytics (Google)**
- **Machine Learning Applications and Chatbots**
- **Social Media Analytics Advanced Text Mining like Sentiment Analysis, Topic Modelling, and Text Summarisation**

Module 9 – Data Science At Scale with Spark

Introduction to Big Data And Spark

- **Apache spark framework, RDDs, Stopgaps in existing computing methodologies**

RDDs

- **RDD persistence, caching, General operations: Transformation, Actions, and Functions.**
- **Concept of Key-Value pair in RDDs, Other pair, two pair RDDs**
- **RDD Lineage, RDD Persistence, WordCount Program Using RDD Concepts**
- **RDD Partitioning & How it Helps Achieve Parallelization**

Advanced Concepts & Spark-Hive

- **Passing Functions to Spark, Spark SQL Architecture, SQLContext in Spark SQL**
- **User-Defined Functions, Data Frames, Interoperating with RDDs**
- **Loading Data through Different Sources, Performance Tuning**
- **Spark-Hive Integration**



**BIG
DATA**

Spark

Module 10 – Deep Learning Using TensorFlow

Artificial Intelligence Basics

- Introduction to keras API and tensorflow

Neural Networks

- Neural networks
- Multi-layered Neural Networks
- Artificial Neural Networks

Deep Learning

- Deep neural networks
- Convolutional Neural Networks
- Recurrent Neural Networks
- GPU in deep learning
- Autoencoders, restricted boltzmann machine



Module 11 – Natural Language Processing Text Mining, Cleaning, and Pre-processing

- **Various Tokenizers, Tokenization, Frequency Distribution, Stemming, POS Tagging, Lemmatization, Bigrams, Trigrams & Ngrams, Lemmatization, Entity Recognition.**

Text classification, NLTK, sentiment analysis, etc

- **Overview of Machine Learning, Words, Term Frequency, Countvectorizer, Inverse Document Frequency, Text conversion, Confusion Matrix, Naive Bayes Classifier.**

Sentence Structure, Sequence Tagging, Sequence Tasks, and Language Modeling

- **Language Modeling, Sequence Tagging, Sequence Tasks, Predicting Sequence of Tags, Syntax Trees, Context-Free Grammars, Chunking, Automatic Paraphrasing of Texts, Chinking.**

AI Chatbots and Recommendations Engine

- **Using the NLP concepts, build a recommendation engine and an AI chatbot assistant using AI.**

Module 12 – Computer Vision

RBM and DBNs & Variational AutoEncoder

- **Introduction rbm and autoencoders**
- **Deploying rbm for deep neural networks, using rbm for collaborative filtering**
- **Autoencoders features and applications of autoencoders.**

Object Detection using Convolutional Neural Net

- **Constructing a convolutional neural network using TensorFlow**
- **Convolutional, dense, and pooling layers of CNNs**
- **Filtering images based on user queries**

Generating images with Neural Style and Working with Deep Generative Models

- **Automated conversation bots leveraging**
- **Generative model, and the sequence to sequence model (lstm).**

Distributed & Parallel Computing for Deep Learning Models

- **Parallel Training, Distributed vs Parallel Computing**
- **Distributed computing in Tensorflow, Introduction to tf.distribute**
- **Distributed training across multiple CPUs, Distributed Training**
- **Distributed training across multiple GPUs, Federated Learning**
- **Parallel computing in Tensorflow**

Reinforcement Learning

- Mapping the human mind with deep neural networks (dnns)
- Several building blocks of artificial neural networks (anns)
- The architecture of dnn and its building blocks
- Reinforcement learning in dnn concepts, various parameters, layers, and optimization algorithms in dnn, and activation functions.

Deploying Deep Learning Models and Beyond

- Understanding model Persistence, Saving and Serializing Models in Keras, Restoring and loading saved models
- Introduction to Tensorflow Serving, Tensorflow Serving Rest, Deploying deep learning models with Docker & Kubernetes, Tensorflow Serving Docker, Tensorflow Deployment Flask.
- Deploying deep learning models in Serverless Environments
- Deploying Model to Sage Maker
- Explain Tensorflow Lite Train and deploy a CNN model with TensorFlow

Module 13 – Generative AI

- **LSTM – What is LSTM?, How does LSTM work, Applications of LSTM, etc.**
- **Transformers – What are transformers, how does a transformer work in deep learning, applications of transformers, types of transformers, encoder-decoded, self-attention, etc.**
- **BERT – Language Models, What is BERT, How does BERT work, how is BERT different from LSTM, applications of BERT, etc.**
- **GPT – What are generative pre-trained models (GPT), how does a GPT work?, real life examples of GPT, etc.**
- **LLM – NLP and Language models, what are LLMs, how does a LLM work, applications of LLM, etc**

Module 14 – Power BI Power BI Basics

- **Introduction to PowerBI, Use cases and BI Tools , Data Warehousing, Power BI components, Power BI Desktop, workflows and reports , Data Extraction with Power BI.**
- **SaaS Connectors, Working with Azure SQL database, Python and R with Power BI**
- **Power Query Editor, Advance Editor, Query Dependency Editor, Data Transformations, Shaping and Combining Data ,M Query and Hierarchies in Power BI.**



DAX

- **Data Modeling and DAX, Time Intelligence Functions, DAX Advanced Features**

Data Visualization with Analytics

- **Slicers, filters, Drill Down Reports**
- **Power BI Query, Q & A and Data Insights**
- **Power BI Settings, Administration and Direct Connectivity**
- **Embedded Power BI API and Power BI Mobile**
- **Power BI Advance and Power BI Premium**

Hands-on Exercise:

Creating a dashboard to depict actionable insights in sales data.

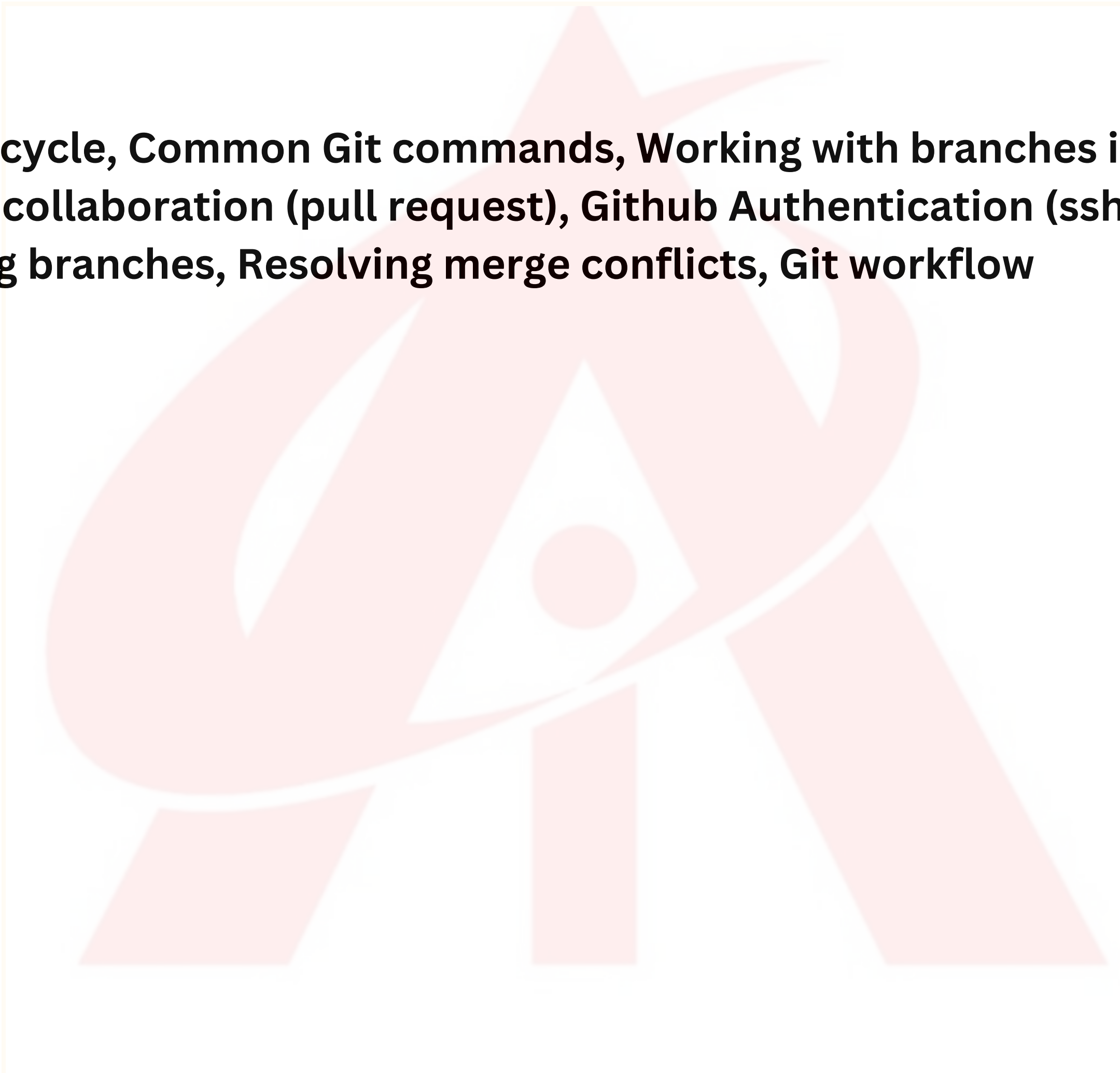
Module 15 – Git

Version Control

- What is version control, types, SVN.

GIT

- Git Lifecycle, Common Git commands, Working with branches in Git
- Github collaboration (pull request), Github Authentication (ssh and Http)
- Merging branches, Resolving merge conflicts, Git workflow



Module 16 – Azure Data Engineering

1. Non-Relational Data Stores and Azure Data Lake Storage

1.1 Document data stores

1.2 Columnar data stores

1.3 Key/value data stores

1.4 Graph data stores

1.5 Time series data stores

1.6 Object data stores

1.7 External index

1.8 Why NoSQL or Non-Relational DB?

1.9 When to Choose NoSQL or Non-Relational DB?

1.10 Azure Data Lake Storage

Definition, Azure Data Lake-Key Components, How it stores data?

Azure Data Lake Storage

Gen2, Why Data Lake? Data Lake Architecture

2. Data Lake and Azure Cosmos DB

2.1 Data Lake Key Concepts

2.2 Azure Cosmos DB

2.3 Why Azure Cosmos DB?

2.4 Azure Blob Storage

2.5 Why Azure Blob Storage?

2.6 Data Partitioning: Horizontal partitioning, vertical partitioning, Functional partitioning

2.7 Why Partitioning Data?

2.8 Consistency Levels in AzureCosmos DB: Semantics of the five-consistency level

3. Relational Data Stores

3.1 Introduction to Relational Data Stores



3.2 Azure SQL Database – Deployment Models, Service Tiers

3.3 Why SQL Database Elastic Pool?

4. Why Azure SQL?

4.1 Azure SQL Security Capabilities

4.2 High-Availability and Azure SQL Database: Standard Availability Model, Premium Availability Model

4.3 Azure Database for MySQL

4.4 Azure Database for PostgreSQL

4.5 Azure Database for MariaDB

4.6 What is PolyBase and Why PolyBase?

4.7 What is Azure Synapse Analytics (formerly SQL DW): SQL Analytics and SQL pool in Azure

Synapse, Key component of a big data solution, SQL Analytics MPP architecture components

5. Azure Batch

5.1 What is Azure Batch?

5.2 Intrinsically Parallel Workloads

5.3 Tightly Coupled Workloads

5.4 Additional Batch Capabilities

5.5 Working of Azure Batch

6. Azure Data Factory

6.1 Flow Process of Data Factory

6.2 Why Azure Data Factory

6.3 Integration Runtime in Azure Data Factory

6.4 Mapping Data Flows

7. Azure Data Bricks

7.1 What is Azure Databricks?

7.2 Azure Spark-based Analytics Platform

7.3 Apache Spark in Azure Databricks

8. Azure Stream Analytics

8.1 Working of Stream Analytics

8.2 Key capabilities and benefits

8.3 Stream Analytics Windowing Functions: Tumbling window, Hopping Window, Sliding Window, Session Window

Module 17 – MLOps

Introduction to MLOps

- **MLOps lifecycle**
- **MLOps pipeline**
- **MLOps Components, Processes, etc**

Deploying Machine Learning Models

- **Introduction to Azure Machine Learning**
- **Deploying Machine Learning Models using Azure**



Module 18 – Data Science Capstone Project

The Data Science capstone project focuses on establishing a strong hold of analyzing a problem

and coming up with solutions based on insights from the data analysis perspective. The capstone project will help you master the following verticals:

- **Extracting, loading and transforming data into usable format to gather insights.**
- **Data manipulation and handling to pre-process the data.**
- **Feature engineering and scaling the data for various problem statements.**
- **Model selection and model building on various classification, regression problems using supervised/unsupervised machine learning algorithms.**
- **Assessment and monitoring of the model created using the machine learning models.**

Module 19 - Business Case Studies

- **Recommendation Engine** – The case study will guide you through various processes and techniques in machine learning to build a recommendation engine that can be used for movie recommendations, restaurant recommendations, book recommendations, etc.
- **Rating Predictions** – This text classification and sentiment analysis case study will guide you towards working with text data and building efficient machine learning models that can predict ratings, sentiments, etc.
- **Census** – Using predictive modeling techniques on the census data, you will be able to create actionable insights for a given population and create machine learning models that will predict or classify various features like total population, user income, etc.
- **Housing** – This real estate case study will guide you towards real world problems, where a culmination of multiple features will guide you towards creating a predictive model to predict housing prices.
- **Object Detection** – A much more advanced yet simple case study that will guide you towards making a machine learning model that can detect objects in real time.
- **Stock Market Analysis** – Using historical stock market data, you will learn about how feature engineering and feature selection can provide you some really helpful and actionable insights for specific stocks.
- **Banking Problem** – A classification problem that predicts consumer behavior based on various features using machine learning models.
- **AI Chatbot** – Using the NLTK python library, you will be able to apply machine learning algorithms and create an AI chatbot.

Module 20 – Excel Analytics

Excel Fundamentals

- **Reading the Data, Referencing in formulas , Name Range, Logical Functions, Conditional Formatting, Advanced Validation, Dynamic Tables in Excel, Sorting and Filtering**
- **Working with Charts in Excel, Pivot Table, Dashboards, Data And File Security**
- **VBA Macros, Ranges and Worksheet in VBA**
- **IF conditions, loops, Debugging, etc.**

Excel For Data Analytics

- **Handling Text Data, Splitting, combining, data imputation on text data, Working with Dates in Excel, Data Conversion, Handling Missing Values, Data Cleaning, Working with Tables in Excel, etc.**

Data Visualization with Excel

- **Charts, Pie charts, Scatter and bubble charts**
- **Bar charts, Column charts, Line charts, Maps**
- **Multiples: A set of charts with the same axes, Matrices, Cards, Tiles**

Excel Power Tools

- **Power Pivot, Power Query and Power View**

Classification Problems using Excel

- **Binary Classification Problems, Confusion Matrix, AUC and ROC curve**
- **Multiple Classification Problems**

Information Measure in Excel

- Probability, Entropy, Dependence

- Mutual Information

Regression Problems Using Excel

- Standardization, Normalization, Probability Distributions

- Inferential Statistics, Hypothesis Testing, ANOVA, Covariance, Correlation

- Linear Regression, Logistic Regression, Error in regression, Information Gain using Regression

Hands-on Exercise:

Classification problem using excel on sales data, and statistical tests on various samples from the population.

