Tech Learniversity

JOB GUARANTEED

PROGRAM

QUALITY EDUCATION FROM QUALITY PEOPLE

Tech Learniversity
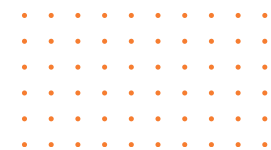
DATA SCIENTIST
PROGRAM

QUALITY EDUCATION FROM QUALITY PEOPLE

DATA SCIENTIST PROGRAM

This Data Scientist Program is an intensive journey through data science foundations, statistical analysis, machine learning, deep learning, and model deployment. Delivered over 400+ hours, it equips students with practical, industry-relevant skills—from Python programming and database queries to advanced AI and MLOps—to solve complex business problems.

Course Overview

Data Scientists are in high demand, as they combine mathematics, programming, and domain knowledge to extract actionable insights. This program covers:
– Core Programming & Data Structures: Python best practices, NumPy, SciPy, and Git.
– Statistics & Probability: From hypothesis testing to regression techniques.
– Data Wrangling & Exploration: SQL, data cleaning with Python (Pandas), feature engineering.
– Machine Learning: Supervised/unsupervised methods, model evaluation, and hyperparameter tuning.
– Deep Learning & Advanced Topics: Neural networks in TensorFlow/PyTorch, NLP, computer vision.
– Deployment & MLOps: Building real-time data solutions with Flask, Docker, cloud environments.
– Ethical AI & Data Privacy: Addressing model bias and compliance (GDPR, CCPA).
– Capstone Project: A real-world project applying all learned concepts end-to-end.

COURSE TYPE

– Designed for a beginner-to-intermediate level.
– Intensifies into advanced machine learning and deep learning concepts, making it robust for those with basic math/programming backgrounds.

Course Objectives

1. Build Strong Foundations: Master Python coding, data structures, and version control.
2. Apply Statistical Methods: Execute descriptive and inferential analyses, including A/B tests.
3. Excel in Data Handling: Acquire, clean, and transform data from diverse sources using SQL, APIs, and web scraping.
4. Master Machine Learning: Implement algorithms for classification, regression, clustering, and dimensionality reduction.
5. Leverage Deep Learning: Construct neural networks for image recognition, NLP, and sophisticated AI solutions.
6. Engineer Production Models: Deploy ML solutions using containers, cloud platforms, monitoring tools, and best MLOps practices.
7. Address Ethical Concerns: Understand biases, fairness, and regulatory frameworks in data science.
8. Demonstrate Full Lifecycle Skills: Present and deploy a real-world capstone project spanning data ingestion to final insights.

DURATION

430 Hours

Requirements

– A computer (Windows, macOS, or Linux) with enough RAM (at least 8 GB recommended)
– Reliable internet connection
– Ability and willingness to install software (Python, SQL databases, IDEs, etc.)

Pre-requisites

– Basic knowledge of linear algebra (e.g., matrices, vectors) and calculus (differentiation, integration)
– Introduction to programming concepts (ideally in Python)
– Understanding of fundamental statistics (mean, variance, distributions) is helpful

TARGET AUDIENCE

– Beginners or early-career professionals targeting Data Science roles
– Software developers aiming to pivot into ML/AI fields
– Statisticians and mathematicians wanting to apply computational techniques
– Data analysts seeking advanced modeling and deployment capabilities
– Professionals from any domain interested in harnessing big data for actionable insights

Career and Future Prospects

Upon completion, graduates can explore roles such as:
– Data Scientist / Research Scientist
– Machine Learning Engineer
– AI Specialist
– Data Analyst (with advanced modeling focus)
– Business Intelligence Specialist (enhanced with data science skills)
With experience, Data Scientists often become team leads, architects, or strategy consultants, shaping AI-driven products and initiatives.

DESIGNATION/TITLE

Common positions in this field include:
– Junior Data Scientist / Associate Data Scientist
– Data Scientist
– Senior Data Scientist
– AI/ML Engineer
– Applied Research Scientist (focus on R&D)

Projects

Hands-on practice and real-world application are critical. The program includes:
1. Data Wrangling & EDA Project
– Collect data from APIs, clean and transform it using Pandas
– Perform exploratory analyses to uncover initial patterns
2. Machine Learning Mini-Projects
– Build classification/regression models (e.g., logistic regression, random forests)
– Tune hyperparameters (e.g., Grid Search, Bayesian optimization)
3. Deep Learning Focus
– Develop a CNN for image classification on a known dataset (e.g., CIFAR-10 or MNIST)
– Explore NLP with RNNs or Transformers for sentiment analysis
4. MLOps / Deployment Task
– Containerize a trained model with Docker
– Serve the model with Flask/Streamlit and demonstrate CI/CD pipelines

PROJECTS

5. Capstone Project
– End-to-end pipeline: data collection, cleaning, model building, deployment, and monitoring
– Presentation of solution architecture, approach, and outcomes to stakeholders

Salary

India : ₹8 LPA – ₹20 LPA
USA : $85,000 – $140,000
Canada : CA$75,000 – CA$120,000
UK : £45,000 – £80,000
Australia : AU$80,000 – AU$120,000

FEATURES

– Comprehensive Curriculum: Covers cutting-edge ML/DL topics along with strong foundations in statistics and programming.
– Practical Approach: Emphasis on coding labs, real data sources, project-based assessments.
– High-End Tools: Hands-on with Python, TensorFlow/PyTorch, Docker, cloud services.
– Collaborative: Regular group projects, peer reviews, and version control workflows.
– Career Support: Resume building, interview practice, network opportunities.

Benefits

– Full Spectrum Expertise: Master everything from data cleaning to advanced AI/ML.
– Industry-Relevant Skills: Projects and proven frameworks used by major tech and data-driven firms.
– Portfolio Development: Capstone project highlighting real-world readiness.
– Scalable Career Path: Move into roles like Senior Data Scientist, ML Engineer, or specialized fields (NLP, CV, etc.).
– Ethical and Compliant: Learn responsible AI practices and data privacy standards.

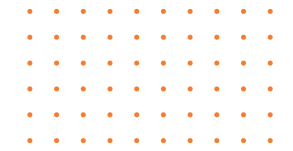## MODULE 1. FOUNDATIONS OF PROGRAMMING AND DATA STRUCTURES

- **Python for Data Science:** Advanced topics including list comprehensions, lambda functions, decorators, generators. Focus on PEP 8 standards and efficient code.
- **Libraries:**
- **NumPy**
- Array creation: np.array(), np.zeros(), np.ones(), np.arange()
- Indexing/slicing: array[0:5], array[:, 1]
- Element-wise operations: +, -, *, /
- Broadcasting rules
- Linear algebra: np.dot(), np.matmul(), np.linalg.inv()
- Statistics: np.mean(), np.median(), np.std()
- Random: np.random.rand(), np.random.randint()
- **SciPy:** (optimization, signal processing).
- **Data Structures:** Efficient use of dictionaries, sets, tuples, and custom classes for data handling (e.g., average O(1) lookup for hash tables).
- **Version Control:** Git and GitHub for collaborative development (branching, merging, pull requests, contributing to open source).

MODULE 2: STATISTICAL INFERENCE AND PROBABILITY

- **Descriptive Statistics:** Measures of central tendency (mean, median, mode), dispersion (variance, standard deviation, IQR), distribution shapes (skewness, kurtosis).
- **Inferential Statistics:** Hypothesis testing (Z-test, T-test, ANOVA, Chi-squared, p-values, significance levels), confidence intervals (95% CI common standard).
- **Probability Theory:** Bayes' Theorem, conditional probability, joint and marginal distributions.
- **Regression Analysis:** Simple and multiple linear regression (R-squared, adjusted R-squared, RMSE), logistic regression for binary classification.
- **Experimental Design:** A/B testing methodologies, power analysis (achieve 80% power at 0.05 significance level).

MODULE 3: DATA COLLECTION, WRANGLING, AND EXPLORATION

- **SQL for Data Science:** Complex queries (JOINs, subqueries, window functions), database design (normalization, indexing for performance: 10-100x faster queries).
- CREATE INDEX
- DROP INDEX, ALTER INDEX
- CREATE VIEW
- DROP VIEW, ALTER VIEW
- CREATE SEQUENCE
- CREATE PROCEDURE / CREATE FUNCTION: Defining parameterized routines
- Control flow (IF/ELSE, WHILE loops, CASE statements)
- ROW_NUMBER(), RANK(), DENSE_RANK(), NTILE()
- LEAD(), LAG(), FIRST_VALUE(), LAST_VALUE()
- Aggregate window functions (SUM() OVER(...), AVG() OVER(...))

- **Data Acquisition:** APIs (RESTful, SOAP, OAuth authentication), web scraping (BeautifulSoup, Scrapy for 1000s of pages/hour), data lakes/warehouses.
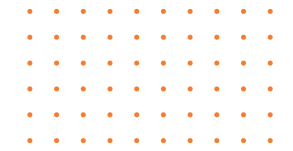
- **Pandas for Data Manipulation:** DataFrames, Series,GroupBy operations, merging, reshaping (pivot, melt), handling missing data (imputation techniques: mean, median, mode, K-NN), outlier detection (IQR, Z-score).
- Create DataFrame: pd.DataFrame()
- Import/export: pd.read_csv(), df.to_csv()
- Data inspection: df.head(), df.info(), df.describe()
- Missing values: df.isnull(), df.fillna(), df.dropna()
- Duplicates: df.drop_duplicates()
- Selection: df['col'], df.loc[], df.iloc[]
- Filtering: df[df['col'] > 10]
- Transformations: df.apply(), df.map(), df.transform()
- Grouping: df.groupby()
- Merge/join: pd.merge(), pd.concat()
- Pivot/melt: df.pivot_table(), pd.melt()
- **Feature Engineering:** Creating new variables from existing ones (e.g., polynomial features, interaction terms, date/time features), encoding categorical variables (one-hot, label, target encoding).
- **Exploratory Data Analysis (EDA):** Univariate, bivariate, multivariate analysis, correlation matrices (Pearson, Spearman coefficients), distribution plots, scatter plots, box plots.
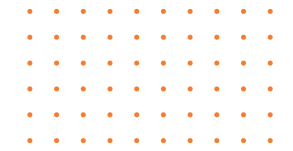
MODULE 4. MACHINE LEARNING FUNDAMENTALS

- **Supervised Learning:**
- **Classification:** k-Nearest Neighbors, Naive Bayes, Support Vector Machines (SVMs), Decision Trees, Random Forests (often 5-10% accuracy boost over single trees), Gradient Boosting (XGBoost, LightGBM, CatBoost: often top performers in Kaggle competitions).
- **Regression:** Ridge, Lasso, Elastic Net regularization.
- **Unsupervised Learning:**
- **Clustering:** K-Means, DBSCAN, Hierarchical Clustering (evaluating with Silhouette score, Elbow method).
- **Dimensionality Reduction:** Principal Component Analysis (PCA: e.g., reducing 100 features to 10 while retaining 95% variance), t-SNE, UMAP.
- **Model Evaluation:** Cross-validation (K-fold), performance metrics (accuracy, precision, recall, F1-score, ROC-AUC for classification; MAE, MSE, R-squared for regression), confusion matrices.
- **Model Selection & Hyperparameter Tuning:** Grid Search, Random Search, Bayesian Optimization (e.g., using Optuna or Hyperopt, can find optimal hyperparameters 2-5x faster).

MODULE 5: DEEP LEARNING AND ADVANCED TOPICS

- **Neural Networks:** Perceptrons, Multi-Layer Perceptrons (MLPs), activation functions (ReLU, sigmoid, tanh), backpropagation.
- **Frameworks:** TensorFlow 2.x, Keras, PyTorch.
- **Convolutional Neural Networks (CNNs):** Image classification (e.g., ResNet, VGG for ImageNet-scale tasks), object detection (YOLO, Faster R-CNN).
- **Recurrent Neural Networks (RNNs):** LSTMs, GRUs for sequence data (NLP, time series).
- **Natural Language Processing (NLP):** Text preprocessing (tokenization, stemming, lemmatization), Word Embeddings (Word2Vec, GloVe), Transformers (BERT, GPT-3 for SOTA NLP tasks).

MODULE 6. DEPLOYMENT, MLOPS, AND ETHICS

- **Model Deployment:** Flask/Streamlit for web apps, Docker for containerization, cloud platforms (AWS Sagemaker, Google AI Platform, Azure ML).
- **MLOps:** CI/CD for ML pipelines, model monitoring (data drift, concept drift), logging, versioning.
- **Big Data Tools:** Apache Spark (processing petabytes of data 100x faster than Hadoop MapReduce), Hadoop, Hive, Kafka for real-time data streams.
- **Data Visualization & Storytelling:** Matplotlib, Seaborn, Plotly, Tableau/Power BI. Effective communication of insights to non-technical stakeholders (e.g., "storytelling with data").
- **Ethics in AI:** Bias in data and models (e.g., racial bias in facial recognition), fairness, accountability, transparency, privacy (GDPR, CCPA compliance).
- **Case Studies & Projects:** End-to-end projects demonstrating problem definition, data acquisition, model building, evaluation, and deployment.
- **Capstone Project:** A comprehensive, real-world project demonstrating proficiency across all syllabus modules, leading to a deployable solution and presentation.

WELCOME TO OUR EDTECH FIRM

Choose Tech Learniversity for quality education from industry experts committed to empowering your success. Embark on a transformative educational journey with Tech Learniversity and unlock your potential for a brighter future. Together, we can create meaningful experiences that inspire growth, collaboration, and success!

Company Description

At Tech Learniversity, we are dedicated to transforming lives through accessible, cutting-edge technology education and personalized learning experiences that empower individuals to excel in the digital era.

ABOUT US

Company Description

Tech Learniversity is a cutting-edge EdTech firm dedicated to revolutionizing global education through personalized learning experiences. We offer innovative training programs, including immersive bootcamps and next-generation AI courses, designed to equip learners with the skills needed to excel in today's competitive job market.

More Information

THANK YOU FOR CHOOSING
TECH LEARNIVERSITY