Machine Learning Workshop with Hugging Face

By Techgyan Technologies

This one-day beginner-friendly workshop is designed for participants with basic or no prior knowledge of Machine Learning (ML). The workshop ensures a step-by-step introduction to ML concepts, focusing on intuitive explanations, hands-on exercises, and real-world applications.

Using Scikit-Learn, TensorFlow, Streamlit, and Hugging Face, participants will build, train, and deploy ML models, including Natural Language Processing (NLP) models, without needing advanced math or coding experience.

The workshop will run ML & Hugging Face side by side, allowing participants to experience structured learning in both traditional ML & cutting-edge AI/NLP models.

Session 1: What is Machine Learning?

- Introduction to AI & ML – How ML differs from traditional programming.
- Real-World Applications – Netflix recommendations, spam filtering, self-driving cars, medical diagnosis.
- Types of ML –
    - Supervised Learning
    - Unsupervised Learning
    - Reinforcement Learning
- ML Workflow –
    - Data Collection → Model Training → Model Testing → Deployment.
- Parallel Hugging Face Activity: Introduction to Hugging Face Model Hub – Exploring pre-trained ML & NLP models.

Session 2: Understanding Data & Feature Engineering

- Types of Data in ML – Categorical vs. Numerical, Structured vs. Unstructured.
- Data Preprocessing – Handling missing values, duplicates, and outliers.
- Feature Engineering – Encoding categorical variables, feature scaling, feature selection.
- Hands-on Activity: Load & clean a simple dataset using Pandas.
- Parallel Hugging Face Activity: Exploring Hugging Face Datasets – Loading and preprocessing text datasets using Hugging Face Datasets library.

Session 3: Splitting Data – Train, Test & Validation

- Why do we split data? – Avoiding overfitting & underfitting.

- Train-Test Split & Validation Set – Understanding Hyperparameter tuning.
- Hands-on Activity: Implement train-test-validation split using Scikit-Learn.
- Parallel Hugging Face Activity: Introduction to Tokenization – Converting text into ML-ready data using Hugging Face Tokenizers.

Session 4: Supervised Learning – Regression & Classification

- What is Supervised Learning? – Concept of input-output mapping using labeled data.
- Regression Models (For Prediction)
    - Linear Regression (Predicting House Prices).
- Classification Models (For Decision Making)
    - Logistic Regression (Spam vs. Not Spam).
- Hands-on Activity: Train a Linear Regression & Logistic Regression model using Scikit-Learn.
- Parallel Hugging Face Activity: Perform Sentiment Analysis using a pre-trained Hugging Face Transformer model.

Session 5: Unsupervised Learning – Clustering & Pattern Recognition

- What is Unsupervised Learning? – ML without labeled data.
- Clustering – Finding groups in data (e.g., Customer Segmentation).
- Understanding K-Means Clustering – Real-life analogy (e.g., organizing groceries).
- Hands-on Activity: Perform K-Means Clustering on a customer dataset.
- Parallel Hugging Face Activity: Perform Named Entity Recognition (NER) using Hugging Face's BERT-based models.

Session 6: How to Evaluate ML Models?

- Why Do We Need Model Evaluation?
    - Understanding Overfitting vs. Underfitting.
    - Importance of train-test split & cross-validation.
- Metrics for Regression – MSE, R^2 Score.
- Metrics for Classification – Accuracy, Precision, Recall, F1-score.
- Hands-on Activity: Evaluate a trained model's accuracy on unseen data.
- Parallel Hugging Face Activity: Fine-tune a Text Classification Model using Hugging Face's Trainer API.

Session 7: Model Improvement & Tuning + Bagging & Boosting

- Improving Model Performance – Feature selection, data balancing, hyperparameter tuning.
- Hyperparameter Tuning Basics – Using GridSearchCV for better predictions.
- Bagging vs. Boosting
    - Bagging (Random Forest) – Improves model stability.

    - Boosting (XGBoost, AdaBoost) – Improves model performance sequentially.
- Hands-on Activity: Implement Random Forest & XGBoost.
- Parallel Hugging Face Activity: Fine-tune a Question-Answering Model using Hugging Face Transformers.

Session 8: Deploying Machine Learning Models

- What is Model Deployment? – Why ML models need to be deployed.
- Different Deployment Methods:
    - Local
    - Cloud
    - API-based deployment
- Hands-on Activity: Deploy an ML model using Streamlit for a web-based prediction app.
- Parallel Hugging Face Activity: Deploy a Transformer-based NLP model on Hugging Face Spaces.

Session 9: ML Hackathon – Build Your Own Model

- Team-based challenge – Solve a real-world problem using ML & Hugging Face.
- Example Projects:
    - Stock Price Prediction – Using Linear Regression.
    - Email Spam Detector – Using Logistic Regression.
    - Text Sentiment Analysis – Using Hugging Face Transformers.
- Teams build, test, and evaluate their models.

Session 10: Wrap-Up & Certificate Distribution

- Recap of Key Learnings.
- Team Presentations.
- Q&A & Next Steps.
- Certificate Distribution.

Who Should Attend This Workshop

Ideal for beginners in technology, software developers, data analysts, product managers, students, entrepreneurs, and educators, this workshop caters to anyone interested in understanding and applying machine learning in real-world scenarios.

Outcome of the One-Day Workshop

Participants will leave with a foundational understanding of machine learning, practical skills in building and deploying ML models using modern tools like Scikit-Learn and TensorFlow, and hands-on experience in Natural Language Processing with Hugging Face. The workshop provides insights into real-world applications, from data handling to model deployment. Attendees will also receive a certificate of completion, expand their professional network, and gain valuable industry-relevant skills.