internStudio

MACHINE LEARNING INTERNSHIP

TASK GUIDE

Turn data into predictions. Build models.
Solve real-world problems.

CORE FOCUS AREAS

- Data Preprocessing
- Exploratory Data Analysis
- Feature Engineering
- Supervised Learning
- Unsupervised Learning
- Model Evaluation
- Model Interpretation
- ML Deployment

INTERNSHIP
Machine Learning

REQUIRED TASKS
Complete Any 2 Projects

CORE TOOLS
Python · Pandas · Scikit-learn

INTERNSHIP INSTRUCTIONS

Your Internship Tasks

Select and complete **any 2 projects** from the task list, developing a **complete machine learning solution** for each — not just a trained model with an accuracy score.

Collect	Clean	Explore	Engineer	Train	Evaluate	Interpret	Deploy

EACH PROJECT SHOULD DEMONSTRATE

- [ ] Data preprocessing and cleaning
- [ ] Exploratory Data Analysis
- [ ] Feature engineering
- [ ] Appropriate ML algorithm selection
- [ ] Training and validation
- [ ] Model evaluation
- [ ] Comparison of multiple models where appropriate
- [ ] Hyperparameter tuning where useful
- [ ] Interpretation of results
- [ ] Clear conclusions and recommendations

EVERY SUBMISSION INCLUDES

- [ ] GitHub repository
- [ ] Jupyter Notebook or Python scripts
- [ ] Dataset / source
- [ ] README
- [ ] Data preprocessing steps
- [ ] EDA visualizations
- [ ] Model training and evaluation
- [ ] Final results and key findings
- [ ] Screenshots where applicable
- [ ] Requirements / dependency file

Suggested technologies: Python · Pandas · NumPy · Matplotlib / Seaborn · Scikit-learn · Jupyter Notebook. Optional: XGBoost / LightGBM for modeling, Streamlit / Flask / FastAPI for deployment.

Submit your completed work here

Once both projects are trained, evaluated, documented and pushed to GitHub, submit your project details through the official submission form.

https://forms.gle/vxL5JE8cvRXXfDBv8

A note on quality. Do not submit a project that only trains a model and prints an accuracy score. You should be able to explain why you chose your algorithm, how well it actually performs, and what the results mean for the underlying problem.

01 Customer Churn Prediction & Retention Intelligence

Build a machine learning system that predicts whether a customer is likely to leave a service.

REQUIREMENTS

- Perform data cleaning and preprocessing
- Analyze customer behavior and churn patterns
- Perform feature engineering
- Identify important factors associated with churn
- Train and compare multiple classification models
- Handle class imbalance where necessary
- Evaluate using Precision, Recall, F1-score, ROC-AUC
- Interpret important features
- Generate actionable retention recommendations
- Optionally build a simple prediction interface

Focus: Classification, feature engineering, model evaluation and business interpretation.

02 Demand & Sales Forecasting System

Build a machine learning solution to predict future product or service demand.

REQUIREMENTS

- Analyze historical sales / demand data
- Clean and prepare time-based data
- Identify trends and seasonal patterns
- Engineer useful time-based features
- Create training and validation datasets appropriately
- Train suitable regression / forecasting models
- Compare model performance
- Evaluate prediction errors
- Visualize actual vs predicted demand
- Recommend inventory or business planning actions

Focus: Regression, time-based features, forecasting, evaluation and business decision-making.

03 Fraud Detection System

Develop a machine learning model capable of identifying potentially fraudulent transactions.

REQUIREMENTS

- Explore transaction patterns
- Clean and preprocess the dataset
- Analyze class imbalance
- Engineer relevant transaction features
- Train and compare classification approaches
- Prioritize fraud-detection metrics over raw accuracy
- Analyze false positives and false negatives
- Identify important predictive features
- Output a fraud-risk prediction

Focus: Classification, imbalanced data, feature engineering, evaluation and model interpretation.

04 Personalized Recommendation Engine

Build a recommendation system that suggests products, movies, courses or other items based on user behavior.

REQUIREMENTS

- Analyze user-item interaction data
- Preprocess interaction / rating information
- Explore user and item behavior
- Implement a recommendation approach
- Generate personalized recommendations
- Handle new users / items where possible
- Evaluate recommendation quality with appropriate metrics
- Compare recommendation strategies
- Create a simple interface to demonstrate recommendations

Focus: Recommendation systems, similarity, user behavior and model evaluation.

05 Employee Attrition Prediction & Workforce Analytics

Build a machine learning solution to predict employee attrition and identify factors associated with turnover.

REQUIREMENTS

- Analyze employee characteristics and workplace factors
- Clean and preprocess the dataset
- Perform exploratory analysis
- Engineer relevant features
- Train and compare classification models
- Evaluate using suitable classification metrics
- Identify the most influential factors
- Analyze different employee segments
- Provide data-driven recommendations to reduce attrition

Focus: Classification, explainability, feature analysis and business insights.

06 Intelligent Image Classification System

Build an image classification model capable of automatically identifying different categories of images.

REQUIREMENTS

- Select an appropriate image dataset
- Prepare and preprocess images
- Split the dataset appropriately
- Train an image classification model
- Evaluate performance using suitable metrics
- Analyze incorrect predictions
- Display sample predictions
- Compare approaches where appropriate
- Save / export the trained model
- Optional: an interface to upload an image and get a prediction

Focus: Computer vision, image preprocessing, classification, evaluation and deployment.

FINAL STEP

Ready to Submit?

Verify all four groups before you send your work in.

DATA

- [ ] Dataset / source included
- [ ] Data cleaning performed
- [ ] Missing values handled
- [ ] Outliers / duplicates addressed where appropriate
- [ ] EDA completed

SUBMISSION

- [ ] GitHub repository
- [ ] Notebook / source code
- [ ] README
- [ ] Dataset / source reference
- [ ] Requirements / dependency file
- [ ] Results / visualizations
- [ ] Screenshots where applicable
- [ ] Deployment / demo link if implemented

MACHINE LEARNING

- [ ] Features engineered
- [ ] Appropriate algorithm selected
- [ ] Train / validation / test strategy used
- [ ] Model trained successfully
- [ ] Multiple models compared where appropriate
- [ ] Hyperparameter tuning considered
- [ ] Appropriate evaluation metrics used

ANALYSIS

- [ ] Model performance explained
- [ ] Important features / results analyzed
- [ ] Errors or incorrect predictions examined
- [ ] Final conclusions documented
- [ ] Practical recommendations provided

Important. Complete any 2 projects. The objective is not simply to achieve a high accuracy score — each project should demonstrate the complete machine learning lifecycle and explain why the model was selected, how well it performs, and what the results mean.

Submission Form

Submit your repository, notebooks, results and key findings here.

https://forms.gle/vxL5JE8cvRXXfDBv8

internStudio

LEARN · BUILD · GROW

www.internstudio.in

team.internstudio@gmail.com