ITSOLERA PVT LTD

AI & Data Science Internship Tasks

Due Date: 20th Jan 2026

Overview

As part of your AI & Data Science Internship at ITSOLERA PVT LTD, you are required to complete **at least 4 out of the 5 tasks** listed below(Task 1 & 2 are mandatory). You are encouraged to complete all of them if you are interested in gaining more hands-on experience.

These tasks are designed to help you develop core skills in data science, such as data exploration, visualization, model building, and performance evaluation using Python. You will use libraries like numpy, pandas, matplotlib, seaborn, and scikit-learn.

Task 1: Data Handling with NumPy & Pandas (Foundations)

Objective

Build a strong foundation in **data loading, cleaning, and manipulation** using NumPy and pandas.

Dataset

Any structured CSV dataset (e.g. Iris, Sales, Student Performance)

Instructions

- Load dataset using pandas
- Inspect data using:
  - `.shape`
  - `.info()`
  - `.columns`
  - `.head()`
- Perform data cleaning:
  - Handle missing values
  - Remove duplicates
- Use NumPy for:
  - Array operations
  - Basic statistics (mean, median, std)

Skills

- NumPy array operations

- pandas data manipulation
- Data cleaning fundamentals

Task 2: Exploring and Visualizing a Simple Dataset

Objective:
Understand how to read, summarize, and visualize a dataset.

Dataset:
Iris Dataset (CSV format, available through seaborn or other open sources)

Instructions:

- Load the dataset using the pandas library.
- Display dataset structure using .shape, .columns, and .head().
- Create basic visualizations:
  - Scatter plot to analyze relationships between variables.
  - Histogram to examine data distribution.
  - Box plot to detect outliers and spread of values.
- Use matplotlib and seaborn for visualizations.

Skills:

- Data loading and inspection using pandas
- Basic data summarization
- Visualization using matplotlib and seaborn

Task 3: Predicting Insurance Claim Amounts

Objective:
Estimate the medical insurance claim amount based on personal data.

Dataset:
Medical Cost Personal Dataset

Instructions:

- Train a Linear Regression model to predict charges.
- Visualize how BMI, age, and smoking status impact insurance charges.
- Evaluate model performance using MAE and RMSE.

Skills:

- Regression modeling
- Feature correlation and visualization
- Error evaluation using MAE and RMSE

Task 4: Credit Risk Prediction

Objective:
Predict whether a loan applicant is likely to default on a loan.

Dataset:
Loan Prediction Dataset (available on Kaggle)

Instructions:

- Handle missing data appropriately.
- Visualize key features such as loan amount, education, and income.
- Train a classification model like Logistic Regression or Decision Tree.
- Evaluate the model using accuracy and a confusion matrix.

Skills:

- Data cleaning and handling missing values
- Exploratory Data Analysis (EDA)
- Binary classification using machine learning
- Model evaluation using confusion matrix and accuracy

Task 5: Customer Churn Prediction (Bank Customers)

Objective:
Identify customers who are likely to leave the bank.

Dataset:
Churn Modelling Dataset

Instructions:

- Clean and prepare the dataset.
- Encode categorical features such as geography and gender.
- Train an ANN Model.
- Analyze important features influencing churn using model weights or permutation importance.

Skills:

- Categorical data encoding (Label Encoding / One-Hot Encoding)
- Deep learning modeling
- Understanding and interpreting feature importance

Submission Requirements (Checklist for Each Task)

To receive credit for each completed task, ensure the following items are included:

1. Jupyter Notebook/ Google Colab or VS Code

- Introduction and problem statement
- Dataset understanding and description
- Data cleaning and preparation
- Exploratory Data Analysis (EDA) with graphs
- Model training and testing
- Evaluation metrics (e.g., accuracy, confusion matrix, MAE, RMSE, etc.)
- Conclusion summarizing key insights

2. Code Quality

- Code should be clean, well-structured, and include comments explaining each step.

3. GitHub Repository

- Create a dedicated GitHub repository for your internship tasks.
- Give the repository a clear and descriptive name.
- Add a README.md file summarizing:
    - The task objective
    - Your approach
    - Results and insights

4. Submission on Google Classroom

- Submit the link to your GitHub repository for each completed task.

Important Note

- **Complete at least 4 out of 5 tasks** before the due

date: **20th Jan 2026**

- You may complete all 5 tasks for more practice and stronger portfolio.

- Ask for help whenever needed.

Task Presentation Requirement

After completing each task, interns are required to **present their work**.

- Short presentation (10-15 minutes)
- Explain problem statement, approach, key visualizations, results, and insights
- Presentations will be conducted **online**
- Interns must be able to clearly explain their own code

Note:

- Tasks without presentation will be considered **incomplete**.
- Missing even one presentation will result in immediate removal from the internship program.