

CSE 322 CLOUD COMPUTING



NAME: ABHISHEK HARSH

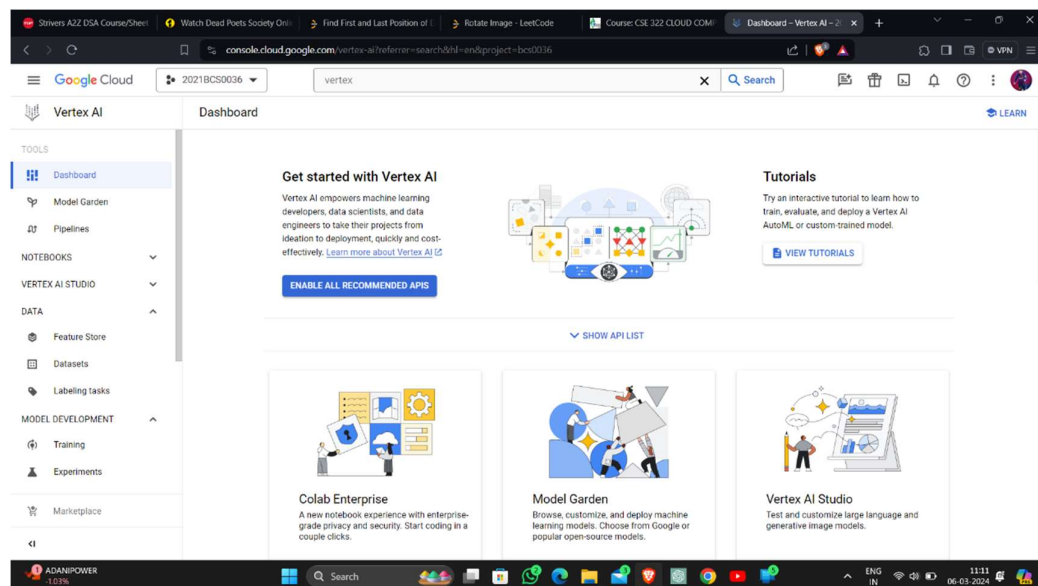
ROLL NO: 2021BCS0036

Lab – 8

AI applications using Machine Learning in GCP

1. Perform the task using tabular Dataset
 - a. Upload and prepare the dataset and perform exploratory data analysis(EDA) to understand the data schema
 - b. Use AutoML to train a model and predict result

Console:



Selecting apis

The screenshot shows the Google Cloud Vertex AI Dashboard. The left sidebar contains navigation options: TOOLS (Dashboard, Model Garden, Pipelines), NOTEBOOKS, VERTEX AI STUDIO, DATA (Feature Store, Datasets, Labeling tasks), MODEL DEVELOPMENT (Training, Experiments), and Marketplace. The main content area is titled 'Get started with Vertex AI' and includes a 'SHOW API LIST' button. A 'Notifications' panel on the right lists enabled services for project 2021BCS0036:

- Enable service: compute.googleapis.com (2021BCS0036)
- Enable service: dataflow.googleapis.com (2021BCS0036)
- Enable service: notebooks.googleapis.com (2021BCS0036)
- Enable service: visionai.googleapis.com (2021BCS0036) - Just now
- Enable service: artifactregistry.googleapis.com (2021BCS0036) - Just now
- Enable service: aiplatform.googleapis.com (2021BCS0036) - Just now
- Enable service: storage-component.googleapis.com (2021BCS0036) - Just now
- Enable service: dataform.googleapis.com (2021BCS0036) - Just now
- Enable service: dataplex.googleapis.com (2021BCS0036) - Just now

The bottom status bar shows the temperature as 89°F (Air: Very Poor) and the time as 11:12 on 06-03-2024.

Create dataset

The screenshot shows the 'Create dataset' page in the Google Cloud Vertex AI console. The 'Dataset name' field is filled with 'abhishek0036'. Below this, the 'Select a data type and objective' section is active, with 'TABULAR' selected. Under 'TABULAR', 'Regression/classification' is chosen as the objective. The 'Region' dropdown is set to 'us-central1 (Iowa)'. The bottom status bar shows the temperature as 89°F (Haze) and the time as 11:12 on 06-03-2024.

Selecting a data source

Vertex AI

TOOLS

- Dashboard
- Model Garden
- Pipelines

NOTEBOOKS

VERTEX AI STUDIO

DATA

- Feature Store
- Datasets**
- Labeling tasks

MODEL DEVELOPMENT

- Training
- Experiments
- Marketplace

abhishek0036

SOURCE ANALYZE

VIEW DATA GUIDE

Select a data source

- CSV file: Can be uploaded from your computer or on Cloud Storage. [Learn more](#)
- BigQuery: Select a table or view from BigQuery. [Learn more](#)

☐ Upload CSV files from your computer

☒ Select CSV files from Cloud Storage

☐ Select a table or view from BigQuery

Select CSV files from Cloud Storage

Enter the Cloud Storage path to one or more CSV files. Data from multiple files will be referenced as one dataset.

Import file path *

gs://spls/cb4455/loan_risk.csv

BROWSE

ADD ANOTHER FILE

What happens next?

The selected CSV file will be associated with your dataset. Making changes to the referenced CSV file will affect the dataset.

CONTINUE

You can build two model types with tabular data. The model type is automatically chosen based on the data type of your target column.

- Regression models predict a numeric value. For example, predicting home prices or consumer spending.
- Classification models predict a category from a fixed number of categories. Examples include predicting whether an email is spam or not, or classes a student might be interested in attending.

Analyzing

Vertex AI

TOOLS

- Dashboard
- Model Garden
- Pipelines

NOTEBOOKS

VERTEX AI STUDIO

DATA

- Feature Store
- Datasets**
- Labeling tasks

MODEL DEVELOPMENT

- Training
- Experiments
- Marketplace

abhishek0036

SOURCE ANALYZE

TRAIN NEW MODEL

Analyze

Dataset location(s)

gs://spls/cb4455/loan_risk.csv

Encryption type

Google-managed

Summary

Total columns: 5

Total rows: -

GENERATE STATISTICS

Filter Enter property name or value

Column name	Missing % (count)	Distinct values
age	-	-
ClientID	-	-
Default	-	-
Income	-	-
loan	-	-

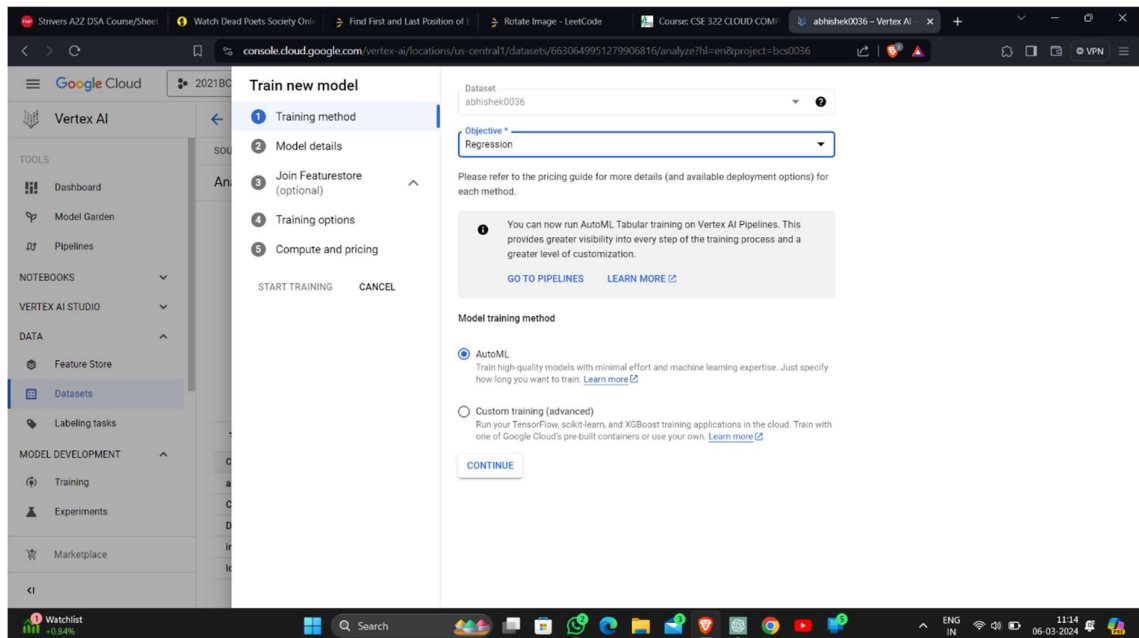
Related resources

Training jobs and models

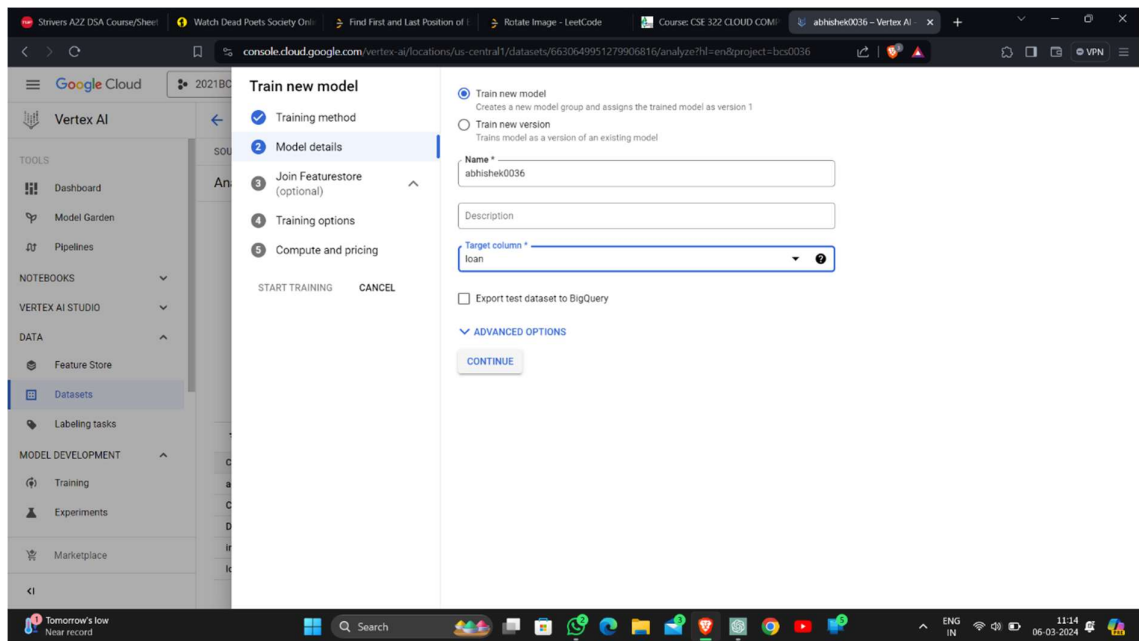
Use this dataset and annotation set to train a new machine learning model with AutoML, or custom code. Selecting AutoML on Pipelines will create a Run on Vertex AI Pipelines. Run information will be found on the [Runs](#) tab under Pipelines.

TRAIN NEW MODEL

Training the new model



Train new model



Train new model(Transformation)

The screenshot shows the 'Train new model' interface in Google Cloud Vertex AI. The left sidebar contains navigation options: TOOLS (Dashboard, Model Garden, Pipelines), NOTEBOOKS, VERTEX AI STUDIO, DATA (Feature Store, Datasets, Labeling tasks), and MODEL DEVELOPMENT (Training, Experiments, Marketplace). The 'Datasets' option is selected. The main panel is titled 'Train new model' and has a progress bar with steps: Training method, Model details, Join Featurestore (optional), Training options (selected), and Compute and pricing. Below the progress bar are 'START TRAINING' and 'CANCEL' buttons. The 'Training options' step shows a table for feature transformations. A warning message at the top states: 'Before continuing, use the Transformation column to review and specify the data types in your dataset. If unspecified, AutoML will try to apply the most relevant transformation option.' The table has columns: Column name, Transformation, Missing % (count), Distinct values, and Correlation w/ target. The rows are: age (Automatic), ClientID (Automatic), Default (Automatic), Income (Automatic), and loan (Target). Below the table, it says 'Total 5 feature columns are included in the training'. There is an 'ADVANCED OPTIONS' section with a 'CONTINUE' button.

Column name	Transformation	Missing % (count)	Distinct values	Correlation w/ target
age	Automatic	-	-	-
ClientID	Automatic	-	-	-
Default	Automatic	-	-	-
Income	Automatic	-	-	-
loan	Target	-	-	-

Setting the budget

The screenshot shows the 'Train new model' interface in Google Cloud Vertex AI, specifically the 'Compute and pricing' step. The left sidebar is the same as in the previous screenshot. The main panel shows the 'Train new model' progress bar with 'Compute and pricing' selected. Below the progress bar are 'START TRAINING' and 'CANCEL' buttons. The 'Compute and pricing' section has a heading 'Enter the maximum number of node hours you want to spend training your model.' and a subtext 'You can train for as little as 1 node hour. You may also be eligible to train with free node hours. [Pricing guide](#)'. There is a 'Budget' input field with the value '1' and a 'Maximum node hours' label. Below this, it says 'Estimated completion: 1 hour' and 'Factors like dataset size and evaluation metrics generation can make training take longer than estimated'. There is a checkbox for 'Enable early stopping' which is checked, with a subtext 'Ends model training when no more improvements can be made and refunds leftover training budget. If early stopping is disabled, training continues until the budget is exhausted.'

View training dataset

The screenshot shows the Google Cloud Vertex AI console for project 'abhishek0036'. The left sidebar contains navigation options: TOOLS (Dashboard, Model Garden, Pipelines), NOTEBOOKS, VERTEX AI STUDIO, DATA (Feature Store, Datasets, Labeling tasks), and MODEL DEVELOPMENT (Training, Experiments, Marketplace). The main content area is titled 'VIEW TRAINING DATASET' and shows a message: 'Training began at Mar 6, 2024, 11:15:20 AM and is still in progress.' Below this, a table lists training details:

Status	Training
Training pipeline ID	441982714925547520
Created	Mar 6, 2024, 11:15:20 AM
Start time	Mar 6, 2024, 11:15:20 AM
Budget (original)	1 node hours
Elapsed time	33 sec
Region	us-central1
Encryption type	Google-managed
Dataset	abhishek0036
Target column	loan
Data split	Randomly assigned (80/10/10)
Column metadata	VIEW DETAILS
Algorithm	AutoML
Objective	Tabular regression
Optimized for	RMSE
Training stage	Data preprocessing

Below the table is a section for 'Training performance'.

Training the model

The screenshot shows the Google Cloud Vertex AI console for project '2021BCS0036'. The left sidebar is the same as in the previous screenshot. The main content area is titled 'Training' and has tabs for 'TRAINING PIPELINES', 'CUSTOM JOBS', 'HYPERPARAMETER TUNING JOBS', 'NAS JOBS', and 'PERSISTENT RESOURCES'. The 'TRAINING PIPELINES' tab is active, showing a table of training pipelines:

Name	ID	Status	Job type	Model type	Duration	Last updated	Created	Ended	Labels
abhishek0036	441982714925547520	Training	Training pipeline	Tabular regression	26 min 30 sec	Mar 6, 2024, 11:15:20 AM	Mar 6, 2024, 11:15:20 AM	-	-
abhishek786	5197625591754391552	Failed	Training pipeline	Tabular classification	9 min 45 sec	Mar 5, 2024, 5:08:16 PM	Mar 5, 2024, 4:58:30 PM	Mar 5, 2024, 5:08:16 PM	-

Training using automl

Google Cloud

2021BCS0036

Search (/) for resources, docs, products, and more

Search

Vertex AI

TOOLS

NOTEBOOKS

VERTEX AI STUDIO

DATA

MODEL DEVELOPMENT

Training

TRAIN NEW MODEL

REFRESH

TRAINING PIPELINES

CUSTOM JOBS

HYPERPARAMETER TUNING JOBS

NAS JOBS

PERSISTENT RESOURCES

Training pipelines are the primary model training workflow in Vertex AI. You can use training pipelines to create an AutoML-trained model or a custom-trained model. For custom-trained models, training pipelines orchestrate custom training jobs and hyperparameter tuning with additional steps like adding a dataset or uploading the model to Vertex AI for prediction serving. [Learn more](#)

Region: us-central1 (Iowa)

Filter: Enter a property name

Name	ID	Status	Job type	Model type	Duration	Last updated	Created	Ended	Labels
abhishek0036	441982714925547520	Finished	Training pipeline	Tabular regression	1 hr 50 min	Mar 6, 2024, 1:06:14 PM	Mar 6, 2024, 11:15:20 AM	Mar 6, 2024, 1:06:14 PM	—
abhishek786	5197625991754391552	Failed	Training pipeline	Tabular classification	9 min 45 sec	Mar 5, 2024, 5:08:16 PM	Mar 5, 2024, 4:58:30 PM	Mar 5, 2024, 5:08:16 PM	—

73°F

Search

ENG IN

14:30

06-03-2024

Predicted Result:

2021BCS0036

Search (/) for resources, docs, products, and more

Search

abhishek0036

Version 1

VIEW DATASET

EXPORT

EVALUATE

DEPLOY & TEST

BATCH PREDICT

VERSION DETAILS

untitled_467921966957252876

COMPARE

CREATE EVALUATION

Target column	MAE	MAPE	RMSE	RMSLE	r ²
loan	1,762.897	148.495	2,244.112	0.837	0.418

Feature importance

Model feature attribution tells you how important each feature is when making a prediction. Attribution values are expressed as a percentage; the higher the percentage, the more strongly that feature impacts a prediction on average. Model feature attribution is expressed using the Sampled Shapley method. [Learn more](#)

