

Lab - 10

BDA : Big Data Analysis

AIM : Mining Complex Types of Data

Name : Ambalia Harshit

Department : Computer Engineering (M.Tech sem II)

Roll no : MT001

Date : 19 March 2024

Title : Using IBM Watson to train data using ML

- Firstly creating New project on IBM Watson

IBM Watson Studio Search in your workspaces Upgrade ?

Projects

Find a project

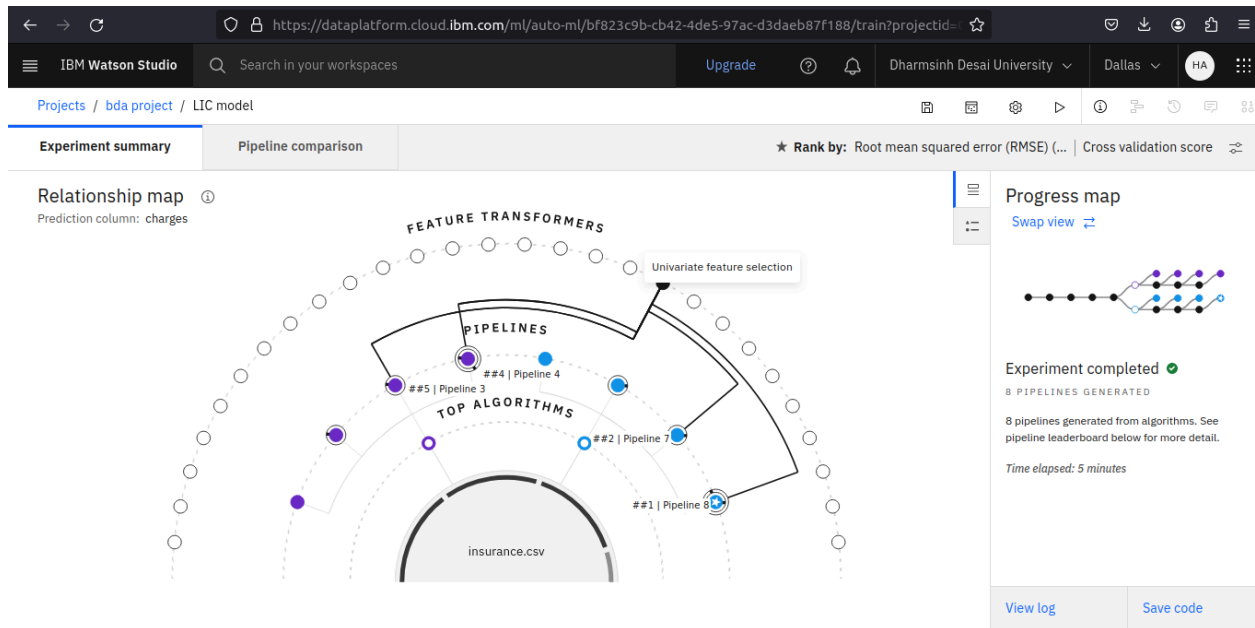
| Name | Date created | Your role |
|-----------------------------|--------------|-----------|
| bda_project | 1 hour ago | Admin |

bda project

- This is our data which will be trained.

| A1 | | age | | | | | | |
|----|-----|--------|--------|----------|--------|-----------|-------------|---|
| | A | B | C | D | E | F | G | H |
| 1 | age | sex | bmi | children | smoker | region | charges | |
| 2 | 19 | female | 27.9 | 0 | yes | southwest | 16884.924 | |
| 3 | 18 | male | 33.77 | 1 | no | southeast | 1725.5523 | |
| 4 | 28 | male | 33 | 3 | no | southeast | 4449.462 | |
| 5 | 33 | male | 22.705 | 0 | no | northwest | 21984.47061 | |
| 6 | 32 | male | 28.88 | 0 | no | northwest | 3866.8552 | |
| 7 | 31 | female | 25.74 | 0 | no | southeast | 3756.6216 | |
| 8 | 46 | female | 33.44 | 1 | no | southeast | 8240.5896 | |
| 9 | 37 | female | 27.74 | 3 | no | northwest | 7281.5056 | |
| 10 | 37 | male | 29.83 | 2 | no | northeast | 6406.4107 | |
| 11 | 60 | female | 25.84 | 0 | no | northwest | 28923.13692 | |
| 12 | 25 | male | 26.22 | 0 | no | northeast | 2721.3208 | |
| 13 | 62 | female | 26.29 | 0 | yes | southeast | 27808.7251 | |
| 14 | 23 | male | 34.4 | 0 | no | southwest | 1826.843 | |
| 15 | 56 | female | 39.82 | 0 | no | southeast | 11090.7178 | |
| 16 | 27 | male | 42.13 | 0 | yes | southeast | 39611.7577 | |
| 17 | 19 | male | 24.6 | 1 | no | southwest | 1837.237 | |
| 18 | 52 | female | 30.78 | 1 | no | northeast | 10797.3362 | |
| 19 | 23 | male | 23.845 | 0 | no | northeast | 2395.17155 | |
| 20 | 56 | male | 40.3 | 0 | no | southwest | 10602.385 | |
| 21 | 30 | male | 35.3 | 0 | yes | southwest | 36837.467 | |
| 22 | 60 | female | 36.005 | 0 | no | northeast | 13228.84695 | |
| 23 | 30 | female | 32.4 | 1 | no | southwest | 4149.736 | |
| 24 | 18 | male | 34.1 | 0 | no | southeast | 1137.011 | |
| 25 | 34 | female | 31.92 | 1 | yes | northeast | 37701.8768 | |

- This platform will use all types of ML methods to train the model by specifying all the kinds of hyper parameters.



- As you can see, it have trained the various types of models by different pipelines. And it is ranked by the accuracy score of all ML models.

| Rank by: Root mean squared error (RMSE) ... Cross validation score | | | | | | | |
|--|---|----------------------------|---------------------------------|--------------------------------------|----------------|------------|-------------------------|
| Rank | ↑ | Name | Algorithm | RMSE (Optimized) Cross Validation | Enhancements | Build time | |
| ★ 1 | | Pipeline 8 | Random Forest Regressor | 4542.283 | HPO-1 FE HPO-2 | 00:00:55 | Save as |
| 2 | | Pipeline 7 | Random Forest Regressor | 4553.868 | HPO-1 FE | 00:00:45 | |
| 3 | | Pipeline 6 | Random Forest Regressor | 4587.601 | HPO-1 | 00:00:04 | |
| 4 | | Pipeline 4 | Snap Boosting Machine Regressor | 4594.936 | HPO-1 FE HPO-2 | 00:00:24 | |
| 5 | | Pipeline 3 | Snap Boosting Machine Regressor | 4594.936 | HPO-1 FE | 00:00:21 | |
| 6 | | Pipeline 2 | Snap Boosting Machine Regressor | 4653.897 | HPO-1 | 00:00:04 | |
| 7 | | Pipeline 1 | Snap Boosting Machine Regressor | 4885.238 | None | 00:00:01 | |

- As we can see Pipeline 8 has the highest accuracy, we can see specifications for that model.

Pipeline details

Pipeline 8

Rank

1

RMSE (Optimized)

3962.685 (Holdout)

Algorithm

Random Forest Regressor

Enhancements

HPO-1 +2

Save as

Model viewer

Model information

Feature summary

Evaluation

Model evaluation

All features

Search feature or transformer names

| Feature name | Transformation | Feature importance |
|--------------|------------------------------|--------------------|
| smoker | None | 70.27% |
| bmi | None | 15.29% |
| age | None | 7.86% |
| NewFeature_0 | featureagglomeration(ALL)[0] | 5.67% |
| children | None | 0.63% |
| NewFeature_1 | featureagglomeration(ALL)[1] | 0.23% |
| region | None | 0.03% |

- We can also download the ipynb code/ source code file which is used to train that model.

← → ↺

https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/e9f7b2ac-02c3-451d-b1cb-c5846cdf4ad3/view. ☆

IBM Watson Studio

Search in your workspaces

Upgrade ?

🔔

Dharmansinh Desai University ▾

Dallas ▾

HA

⋮

Projects / bda project / LIC model - P8 Random Forest Reg

📄 🔗 📄 ▾

Download

🔗 AutoAI Part of IBM Watson® Studio

Pipeline notebook

Pipeline 8 Notebook - AutoAI Notebook v1.22.7

Consider these tips for working with an auto-generated notebook:

- Notebook code generated using AutoAI will execute successfully. If you modify the notebook, we cannot guarantee it will run successfully.
- This pipeline is optimized for the original data set. The pipeline might fail or produce sub-optimal results if used with different data. If you want to use a different data set, consider retraining the AutoAI experiment to generate a new pipeline. For more information, see [Cloud Platform](#).
- Before modifying the pipeline or trying to re-fit the pipeline, consider that the code converts dataframes to numpy arrays before fitting the pipeline (a current restriction of the preprocessor pipeline).

Notebook content

This notebook contains a Scikit-learn representation of AutoAI pipeline. This notebook introduces commands for retrieving data, training the model, and testing the model.

Some familiarity with Python is helpful. This notebook uses Python 3.10 and scikit-learn 1.1.1.

Notebook goals

<https://dataplatform.cloud.ibm.com/analytics/notebooks/v2/e9f7b2ac-02c3-451d-b1cb-c5846cdf4ad3/view?projectId=08b7fabd-0af8-4a47-b367-07d3d67ccefb&context=cpdaas#>

- Now as the model has been created, we can deploy this model online as so that public can use this model.
 - So now we are creating the new deployment for this model.

Create a deployment

Associated asset
LIC model - P8 Random Forest Regressor - Model

Deployment type

Online
Run the model on data in real-time, as data is received by a web service.

Batch
Run the model against data as a batch process.

Name
bda deploy

Serving name ⓘ

CancelCreate

- So now we have created deployment named 'bda deploy'.

Deployments / BDA / LIC model - P8 Random Forest Regressor - Model

Deployments

Model details

🔍 Search

New deployment

| Name | Type | Status | Tags | Last modified | |
|------------|--------|----------|------|---|--|
| bda deploy | Online | Deployed | | 47 seconds ago Harshit Ambalia (You) | |

Online deployment ready

The online deployment [bda deploy](#) in space [BDA](#) is ready to accept requests

Today 11:06 PM

Created

Apr 15, 2024, 11:03 PM

Type

wml-hybrid_0.1

Model ID

f7f765a4-2b73-4ce9-9176-e780af...

Software specification

[hybrid_0.1](#)

Hybrid pipeline software specifications

[autoai-kb_rt23.1-py3.10](#)

Description

No description provided.

Tags

Add tags to make assets easier to find

- We can use various language end points to access this model.

Public endpoint

`https://us-south.ml.cloud.ibm.com/ml/v4/deployments/harshitambalia/predictions?version=2021-05-01`

Learn more about the 2021-05-01 version query parameter

Code snippets

cURL | Java | JavaScript | Python | Scala

```
# NOTE: you must set $API_KEY below using information retrieved from your IBM Cloud account (https://dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/ml-authenticate)
curl --insecure -X POST --header "Content-Type: application/x-www-form-urlencoded" --header "Accept: \
application/json" --data-urlencode "grant_type=urn:ibm:params:oauth:grant-type:apikey" \
--data-urlencode "apikey=$API_KEY" "https://iam.cloud.ibm.com/identity/token"

# the above CURL request will return an auth token that you will use as $IAM_TOKEN in the scoring request below
# TODO: manually define and pass values to be scored below
```

- To access this end points, we have to get the API keys, so now we are creating the API key to access our model through public network.

API keys

Create, view, and work with API keys that you have access to manage. IBM Cloud API keys are associated with a user's identity and can be used to access cloud platform and classic infrastructure APIs, depending on the access that is assigned to the user. The following table displays a list of API keys created in this account. [Learn more.](#)

Looking for more options to manage API Keys? Try IBM Cloud® Secrets Manager for creating and leasing API keys dynamically and storing them securely in your own dedicated instance.

View: My IBM Cloud API keys

API keys associated with a user's identity have the same access that the user is assigned across all accounts. To update the access for an API key, assign or remove access for the user.

| Status | Name | Description | Date created |
|--------|---------|-------------|---------------------|
| | Harshit | | 15-4-2024 17:40 GMT |

Items per page: 25 | 1-25 items | Page 1

- We can also use 'Resource usage' directly by accessing the feature provided.

BDA

Overview | Assets | Deployments | Jobs | **Manage**

Space

General | Access control | Environments | **Resource usage**

Resource usage

Usage summary

For this month in this space

0.22 CUH

[Manage service instances](#)

Usage by type

AutoAI online

y-value: AutoAI online ...

x-value: 0.007

Group: AutoAI online ...

0 0.001 0.002 0.003 0.004 0.005 0.006 0.007

- If we go on assets of the project we have created, we can see all the assets that has been used in this project.

The screenshot shows the IBM Watson Studio interface. At the top, there's a navigation bar with 'IBM Watson Studio', a search bar, and user information. Below this, the 'bda project' is selected. The 'Assets' tab is active, showing a list of 4 assets. On the left, there's a sidebar with 'Asset types' including Data, Experiments, Notebooks, and Models, each with a count of 1. The main area displays a table of assets:

| <input type="checkbox"/> | Name | Last modified |
|--------------------------|--|---------------------------------------|
| <input type="checkbox"/> | LIC model - P8 Random Forest Regress... Model | 35 minutes ago Modified by Service |
| <input type="checkbox"/> | LIC model - P8 Random Forest Regresso... Notebook | 37 minutes ago Modified by you |
| <input type="checkbox"/> | LIC model AutoAI experiment | 48 minutes ago Modified by you |
| <input type="checkbox"/> | insurance.csv CSV | 59 minutes ago Modified by you |

- We can use the inbuilt testing feature provided by itself. As we can see, it is asking for the data on which our model will try to generate the output.

The screenshot shows the IBM Watson Studio interface for the 'bda deploy' deployment. The 'Test' tab is active, showing a section for 'Enter input data'. Below this, there's a table with columns: age (double), sex (other), bmi (double), children (double), smoker (other), and region (other). The first row contains data: 19, female, 27.9, 0, yes, southwest. The second and third rows are empty.

| | age (double) | sex (other) | bmi (double) | children (double) | smoker (other) | region (other) |
|---|--------------|-------------|--------------|-------------------|----------------|----------------|
| 1 | 19 | female | 27.9 | 0 | yes | southwest |
| 2 | | | | | | |
| 3 | | | | | | |

- After providing the input, we can predict the output generated by this model for given input.

