

Project Kickoff Checklist

- **Project Name:** Predicting Life Expectancy Using Machine Learning
- **Kickoff Date:** 11 June 2020

- **Team Members:**

I am Amey Sawant, a Machine Learning and Deep Learning Learning enthusiast who is always eager to learn new things.

- **Project Background:**

We expect human lifespan to rise as there are lot of medical breakthroughs happening nowadays. We have lot of systems to predict disease, chronic conditions and other maladies to increase human lifespan, but predicting it would be one of the bases of all the system.

- **Identifying stakeholders:**

Healthcare industries would be one of the biggest stakeholders who can sponsor and get benefited from such project. Predicting life expectancy and performing operations and other things depending on that would give optimal results.

- **Review Project Objectives:**

1. Objectives:

To create a Regression Machine Learning model which leverages historical data to predict insights into the future. This problem statement is aimed at predicting Life Expectancy rate of a country given various features.

2. Deliverables:

This project delivers a cloud based Machine Learning solution for Life Expectancy Prediction.

3. Assumptions:

Since we are using a Regression model, its assumed that the features affect the lifespan linearly.

- **Review team member roles & responsibilities:**

Having a single person team, all the project work will be done by me, advisory being the Hemant Kuma Gahlot sir.

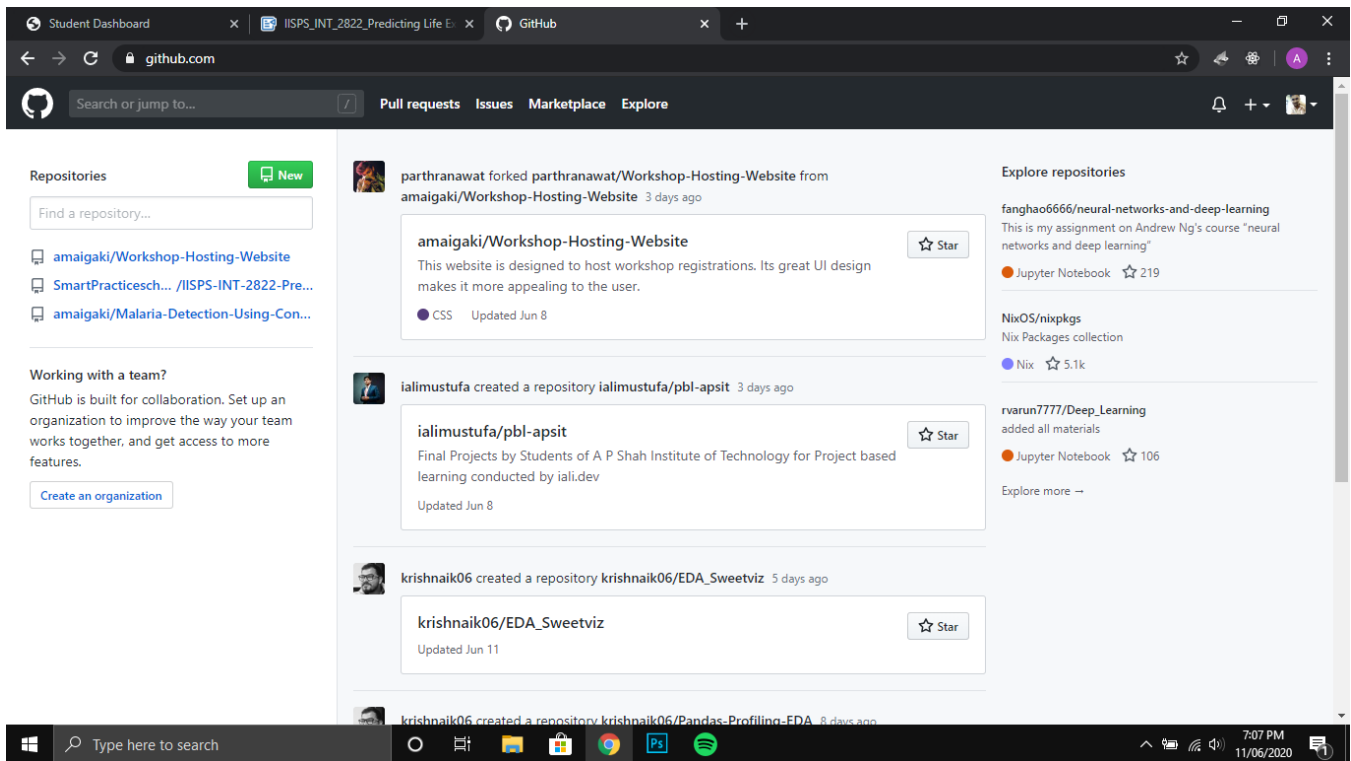
- **Review other potential issues, risks, questions and concerns:**

If not provided with ample amount of data, the model might not predict the life expectancy very accurately. The cloud based model might add up some restrictions based on space and computing in case of larger data and usage.

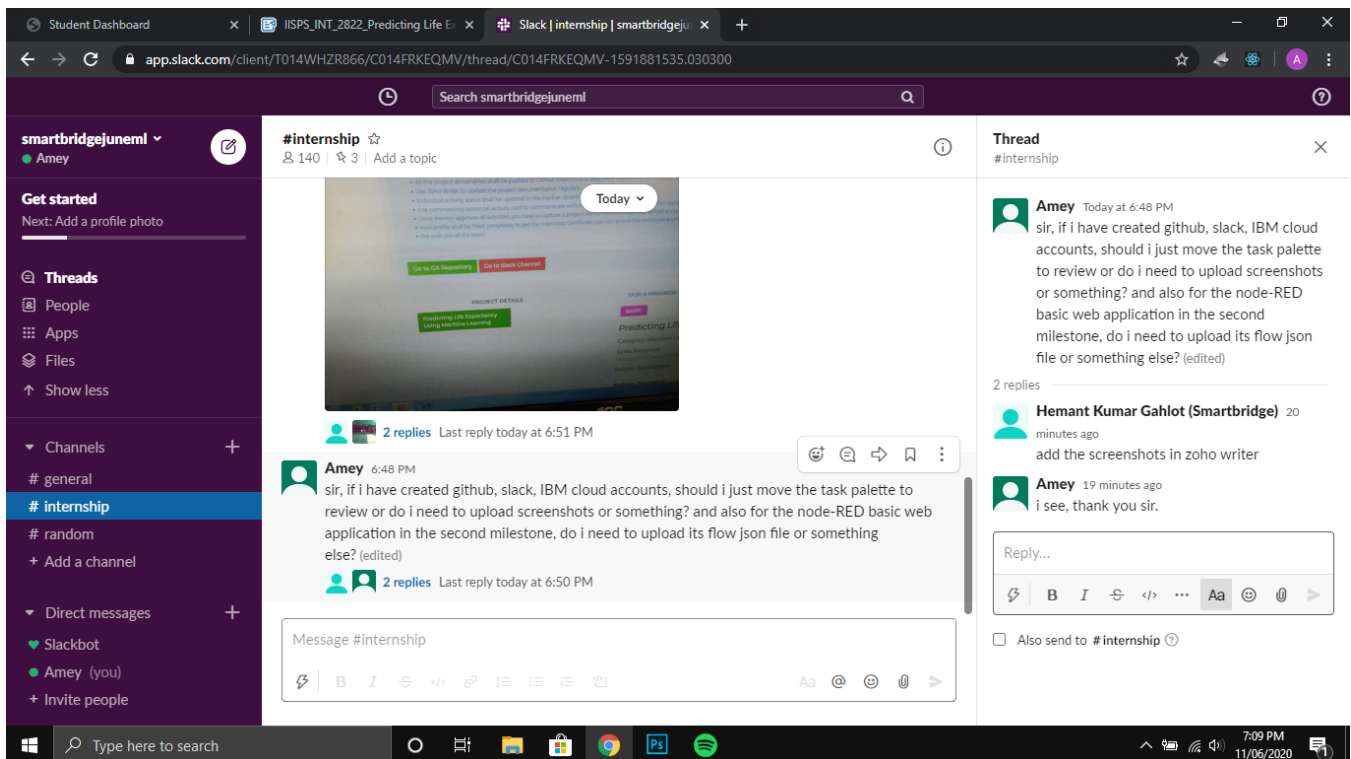
- **Identify next steps and timing:**

Each step of the project will be done under supervision of advisory. The results and task completions will be informed through kanban dashboard.

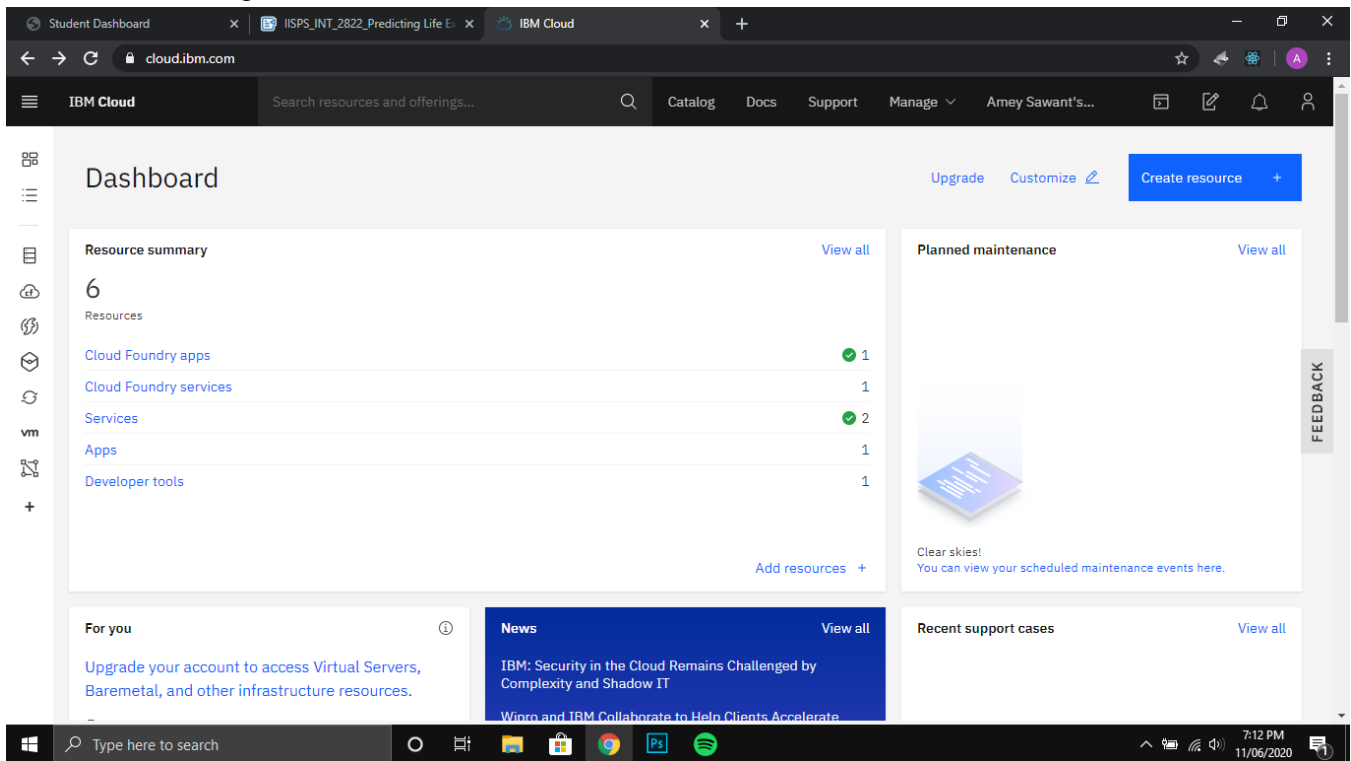
• Created Github Account:



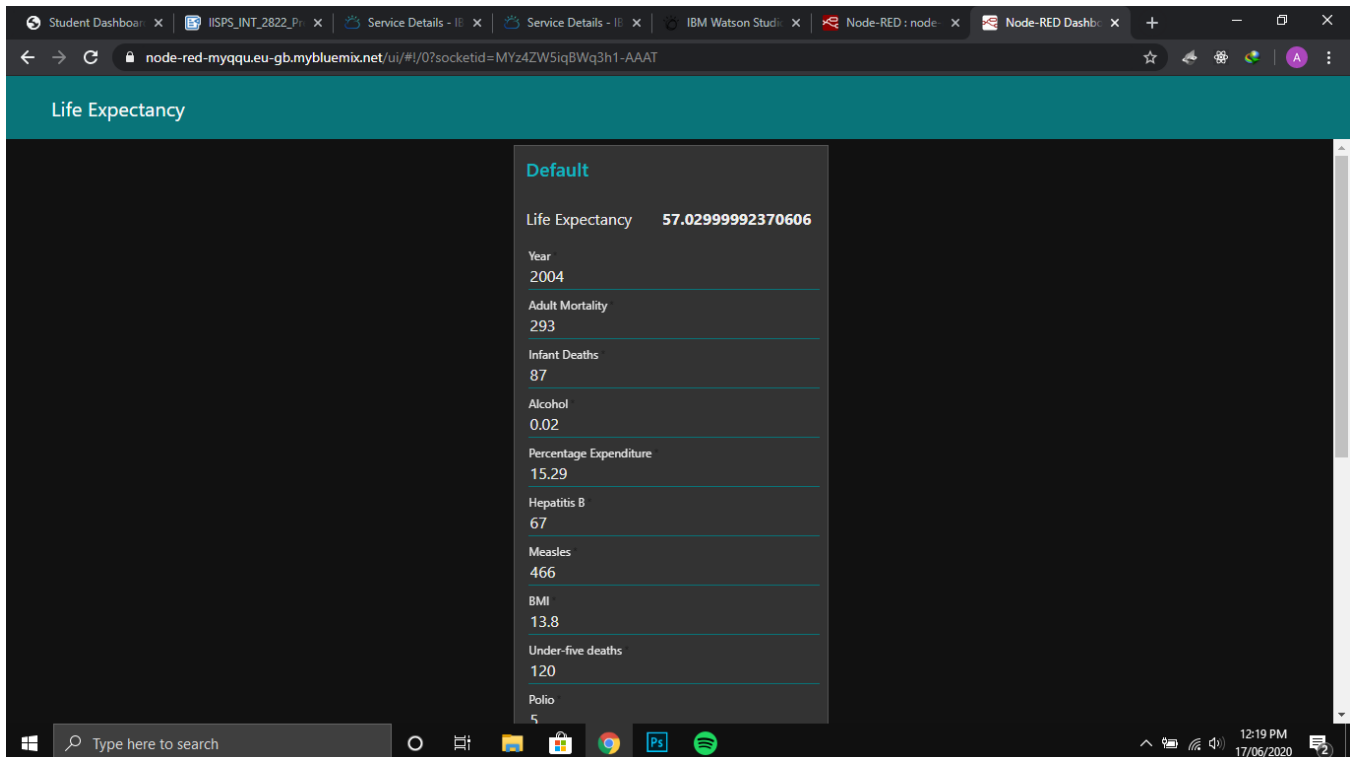
• Joined Slack Channel:



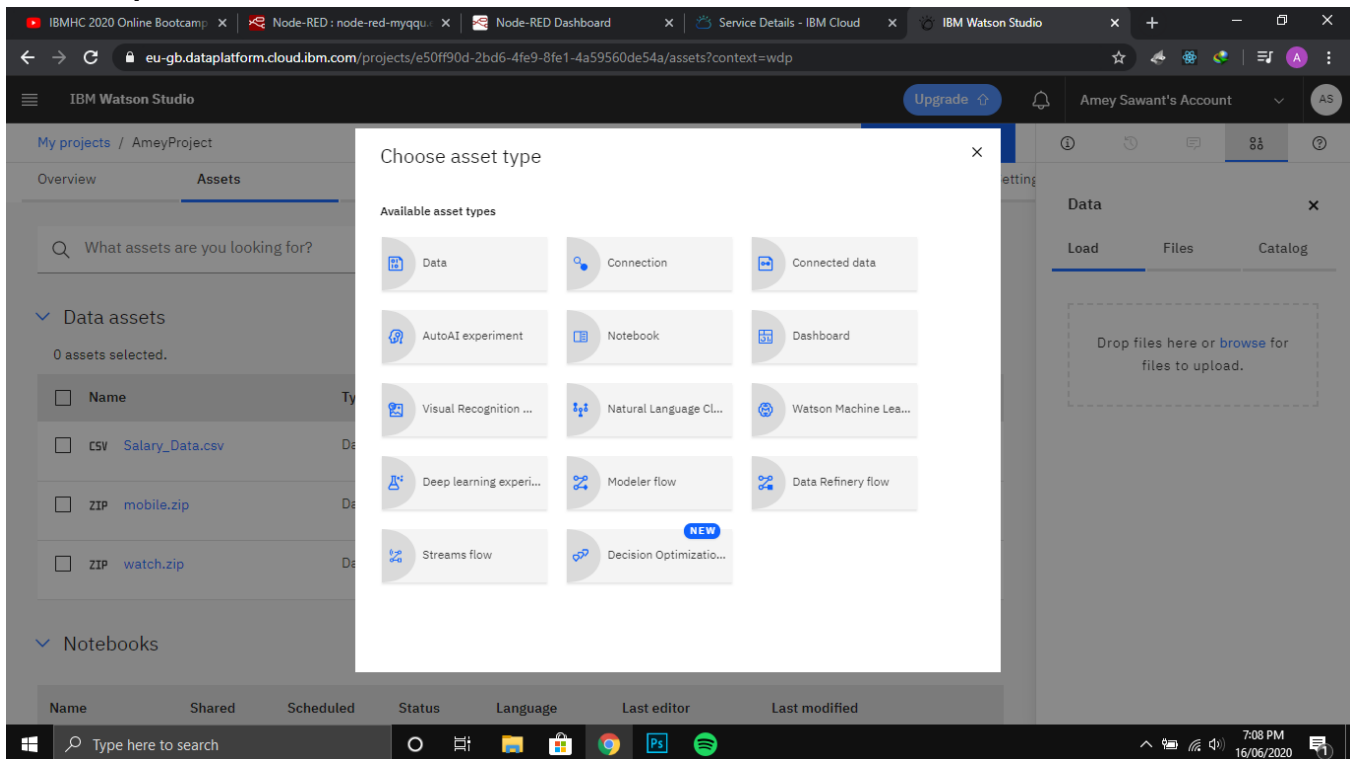
- Created & got started with IBM cloud:



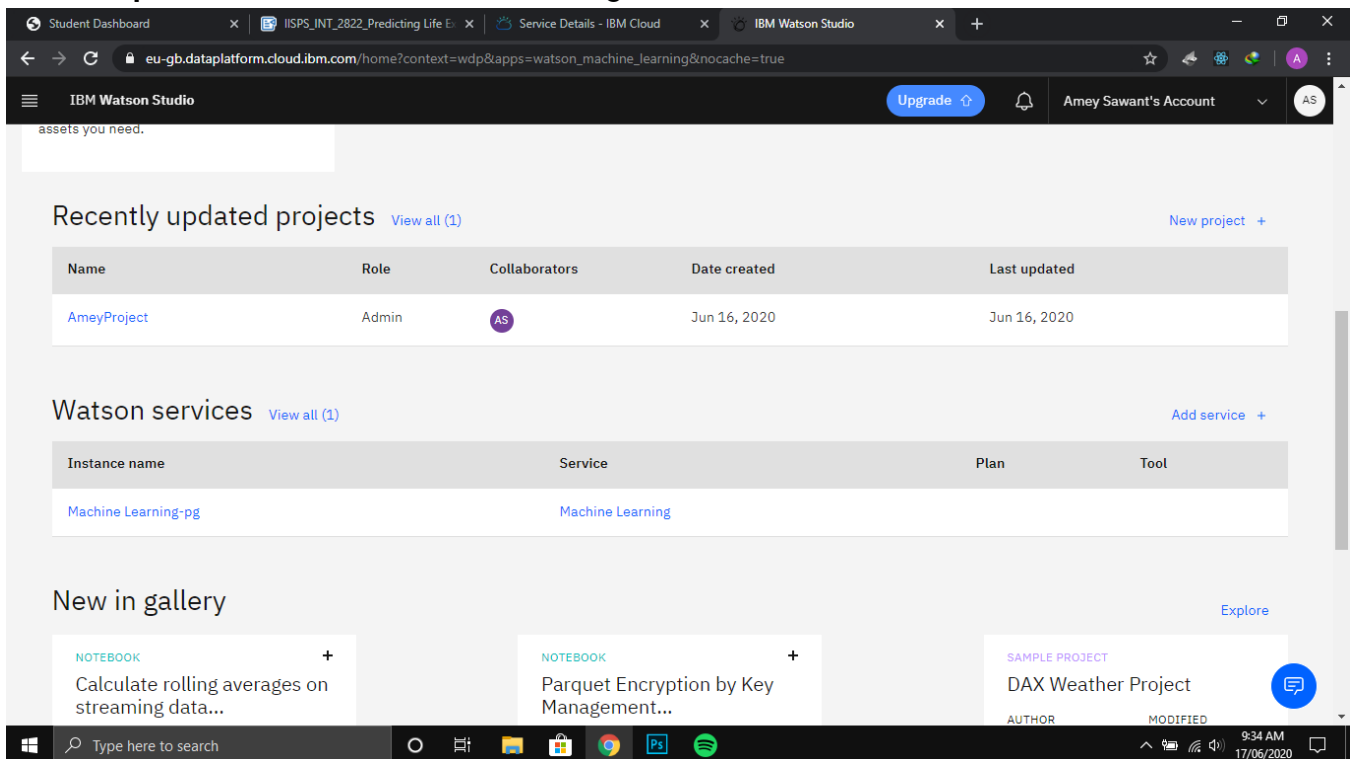
- Created a Starter webpage with node-RED:



- Explored Watson use cases and services:



- Explored IBM watson Machine Learning:



The screenshot shows the IBM Watson Studio web interface. The browser tabs include 'Student Dashboard', 'IISPS_INT_2822_Predicting Life E...', 'Service Details - IBM Cloud', and 'IBM Watson Studio'. The URL is 'eu-gb.dataplatform.cloud.ibm.com/projects/e50ff90d-2bd6-4fe9-8fe1-4a59560de54a/assets?context=wdp'. The user is logged in as 'Amey Sawant's Account'.

The main content area is titled 'My projects / AmeyProject'. It features a table of models and sections for Notebooks and Deep learning experiments.

Name	Status	Model type	Last modified
Salary Predict	Completed	Regression	Jun 16, 2020, 07:23 PM

Notebooks

Name	Shared	Scheduled	Status	Language	Last editor	Last modified
Life Expectancy Prediction				Python 3.6	Amey Sawant	Jun 16, 2020

Deep learning experiments

You don't have any Deep learning experiments yet

The right sidebar shows a 'Data' section with tabs for 'Load', 'Files', and 'Catalog'. A message says: 'Drop files here or browse for files to upload.'

- Built my own models using Watson Studio Machine Learning Services:

This screenshot shows the same IBM Watson Studio interface, but with the 'Models' section expanded. The 'Deep learning experiments' section is now empty, displaying the message 'You don't have any Deep learning experiments yet'.

Models

Visual Recognition models

Name	Model ID	Model type	Last modified
Amey Model	AmeyModel_1416257232	Classification	Jun 16, 2020, 12:43 PM

Watson Machine Learning model

Name	Type	Runtime	Last modified
Salary Predict - P4 ExtraTreesRegressorEstimator	wml-hybrid_0.1	hybrid_0.1	Jun 16, 2020
Salary_Data	scikit-learn-0.20	python-3.6	Jun 16, 2020

The right sidebar remains the same, showing the 'Data' section with a message: 'Drop files here or browse for files to upload.'

- Automated ML model using AutoAI:

IBM Watson Studio

My projects / AmeyProject

You don't have any Deep learning experiments yet

Visual Recognition models

Name	Model ID	Model type	Last modified
Amey Model	AmeyModel_1416257232	Classification	Jun 16, 2020, 12:43 PM

Watson Machine Learning model

Name	Type	Runtime	Last modified
Salary Predict - P4 ExtraTreesRegressorEstimator	wml-hybrid_0.1	hybrid_0.1	Jun 16, 2020
Salary_Data	scikit-learn-0.20	python-3.6	Jun 16, 2020

Data

Load Files Catalog

Drop files here or browse for files to upload.

IBM Watson Studio

My projects / AmeyProject / Salary Predict

Experiment summary Pipeline comparison Rank by: Root mean squared err... Score: Cross validation Holdout

Read dataset Split holdout data Read training data Preprocessing Model selection

Extra Trees Regressor Hyperparameter optimization Feature engineering Hyperparameter optimization

Selected algorithm Hyperparameter optimization Feature engineering Hyperparameter optimization

Feature engineering
EXTRA TREES REGRESSOR
Started feature engineering for pipeline P3
Time elapsed: 117 seconds

Pipeline leaderboard

IBM Watson Studio interface showing a project named "Salary Predict". The interface includes a top navigation bar with "My projects / AmeyProject / Salary Predict" and a "Cross validation" button. Below this is a table titled "Experiment summary" with columns: Rank, Name, Algorithm, RMSE (Optimized), Enhancements, and Build time. The table lists 8 pipelines, with Pipeline 4 having the lowest RMSE (4954.377) and Pipeline 5 having the highest (7152.198). The bottom of the image shows a Windows taskbar with various application icons and a system clock indicating 7:23 PM on 16/06/2020.

Rank	Name	Algorithm	RMSE (Optimized)	Enhancements	Build time
1	Pipeline 4	Extra Trees Regressor	4954.377	HPO-1 FE HPO-2	00:00:13
2	Pipeline 7	Random Forest Regressor	5177.478	HPO-1 FE	00:00:44
3	Pipeline 8	Random Forest Regressor	5177.478	HPO-1 FE HPO-2	00:00:23
4	Pipeline 3	Extra Trees Regressor	5216.874	HPO-1 FE	00:00:13
5	Pipeline 2	Extra Trees Regressor	5436.723	HPO-1	00:00:08
6	Pipeline 1	Extra Trees Regressor	6147.730	None	00:00:01
7	Pipeline 6	Random Forest Regressor	6454.767	HPO-1	00:00:07
8	Pipeline 5	Random Forest Regressor	7152.198	None	00:00:01

Collected Dataset for the project:

Life_Expectancy - Excel

Country	Year	Status	Life expectancy	Adult Mortality	infant deaths	Alcohol	percentage expendi	Hepatitis I	Measles	BMI	under-five	Polio	Total exp	Diphtheri	HIV/AIDS	GDP	Po
Afghanistan	2015	Developing	65	263	62	0.01	71.27962362	65	1154	19.1	83	6	8.16	65	0.1	584.2592	33
Afghanistan	2014	Developing	59.9	271	64	0.01	73.52358168	62	492	18.6	86	58	8.18	62	0.1	612.6965	33
Afghanistan	2013	Developing	59.9	268	66	0.01	73.21924272	64	430	18.1	89	62	8.13	64	0.1	631.745	31
Afghanistan	2012	Developing	59.5	272	69	0.01	78.1842153	67	2787	17.6	93	67	8.52	67	0.1	669.959	33
Afghanistan	2011	Developing	59.2	275	71	0.01	7.097108703	68	3013	17.2	97	68	7.87	68	0.1	63.53723	2
Afghanistan	2010	Developing	58.8	279	74	0.01	79.67936736	66	1989	16.7	102	66	9.2	66	0.1	553.3289	2
Afghanistan	2009	Developing	58.6	281	77	0.01	56.76221682	63	2861	16.2	106	63	9.42	63	0.1	445.8933	2
Afghanistan	2008	Developing	58.1	287	80	0.03	25.87392536	64	1599	15.7	110	64	8.33	64	0.1	373.3611	2
Afghanistan	2007	Developing	57.5	295	82	0.02	10.91015998	63	1141	15.2	113	63	6.73	63	0.1	369.8358	26
Afghanistan	2006	Developing	57.3	295	84	0.03	17.17151751	64	1990	14.7	116	58	7.43	58	0.1	272.5638	2
Afghanistan	2005	Developing	57.3	291	85	0.02	1.388647732	66	1296	14.2	118	58	8.7	58	0.1	25.29413	2
Afghanistan	2004	Developing	57	293	87	0.02	15.29606643	67	466	13.8	120	5	8.79	5	0.1	219.1414	24
Afghanistan	2003	Developing	56.7	295	87	0.01	11.08905273	65	798	13.4	122	41	8.82	41	0.1	198.7285	2
Afghanistan	2002	Developing	56.2	3	88	0.01	16.88735091	64	2486	13	122	36	7.76	36	0.1	187.846	21
Afghanistan	2001	Developing	55.3	316	88	0.01	10.5747282	63	8762	12.6	122	35	7.8	33	0.1	117.497	2
Afghanistan	2000	Developing	54.8	321	88	0.01	10.42496	62	6532	12.2	122	24	8.2	24	0.1	114.56	2
Albania	2015	Developing	77.8	74	0	4.6	364.9752287	99	0	58	0	99	6	99	0.1	3954.228	33
Albania	2014	Developing	77.5	8	0	4.51	428.7490668	98	0	57.2	1	98	5.88	98	0.1	4575.764	33
Albania	2013	Developing	77.2	84	0	4.76	430.8769785	99	0	56.5	1	99	5.66	99	0.1	4414.723	33
Albania	2012	Developing	76.9	86	0	5.14	412.4433563	99	9	55.8	1	99	5.59	99	0.1	4247.614	33
Albania	2011	Developing	76.6	88	0	5.37	437.0621	99	28	55.1	1	99	5.71	99	0.1	4437.179	33
Albania	2010	Developing	76.2	91	1	5.28	41.82275719	99	10	54.3	1	99	5.34	99	0.1	494.3588	33

- Created necessary IBM cloud service:

The screenshot shows the IBM Cloud Application Details page for a service named 'Node RED Life Expectancy'. The application is in a 'Running' state. The left sidebar contains navigation links: Getting started, Overview (selected), Runtime, Connections, Logs, API Management, Autoscaling, and Availability Monitoring. The main content area displays several metrics:

- Overview:** Shows '100%' with the text '1/1 instance(s) are running'. Below this is a slider for 'MB memory per instance' ranging from 0 to 256, with a current value of 128.
- Runtime cost:** Displays '\$ 0.00' for 'Current charges for billing period' and '\$ 0.00' for 'Estimated total for billing period Jun 1, 2020 - Jun 30, 2020'.
- Connections (1):** Lists a single connection: 'node-red-myqqu-cloudant-1591874718743-71817'.
- Memory Usage:** A donut chart shows '128 Total MB allocation' with '128 MB still available'. The chart is divided into 'Used' (purple) and 'Free' (grey) segments.

The bottom of the screen shows a Windows taskbar with various application icons and a system clock indicating 10:11 AM on 17/06/2020.

- Created Watson Studio Project:

The screenshot displays the IBM Watson Studio 'My projects' page. The top navigation bar includes 'IBM Watson Studio', an 'Upgrade' button, and a user profile for 'Amey Sawant's Account'. Below the navigation bar, there is a search bar with the placeholder text 'Which project are you looking for?' and a dropdown menu for 'All my projects'. A 'New project +' button is located in the top right corner. The main content area features a table with project details:

Name	Role	Storage	Collaborators	Creator	Date created
AmeyProject	Admin	COS	AS	Amey Sawant	16 Jun 2020

Below the table is a 'Show more' button. The bottom of the screen shows a Windows taskbar with various application icons and a system clock indicating 10:23 AM on 17/06/2020.

- **Configured Watson Studio:**

The screenshot shows the IBM Watson Studio web interface. The top navigation bar includes 'My projects / AmeyProject', 'Launch IDE', and 'Add to project'. The main content area is divided into tabs: Overview, Assets (selected), Environments, Jobs, Deployments, Access Control, and Settings. A search bar at the top of the Assets tab asks 'What assets are you looking for?'. Below it, the 'Data assets' section shows '0 assets selected.' and a table of existing assets.

<input type="checkbox"/>	Name	Type	Created by	Last modified
<input type="checkbox"/>	CSV Life_Expectancy.csv	Data Asset	Amey Sawant	Jun 17, 2020, 10:24 AM
<input type="checkbox"/>	CSV Salary_Data.csv	Data Asset	Amey Sawant	Jun 16, 2020, 05:36 PM
<input type="checkbox"/>	ZIP mobile.zip	Data Asset	Amey Sawant	Jun 16, 2020, 12:03 PM
<input type="checkbox"/>	ZIP watch.zip	Data Asset	Amey Sawant	Jun 16, 2020, 12:03 PM

Below the table, there is a section for 'AutoAI experiments' with a 'New AutoAI experiment' button. On the right side, a 'Data' panel is open, showing options to 'Load', 'Files', or 'Catalog' data, with a message: 'Drop files here or browse for files to upload.'

- **Created Machine Learning Services:**

The screenshot shows the IBM Cloud console interface. The top navigation bar includes 'Resource list /', 'Machine Learning-pg' (with a green 'Active' status), 'Add tags', 'Details', and 'Actions...'. The left sidebar has a 'Manage' section with 'Service credentials' selected. The main content area shows 'Service credentials' with a description: 'You can generate a new set of credentials for cases where you want to manually connect an app or external consumer to an IBM Cloud™ service. [Learn more](#)'. Below this is a search bar and a table of credentials.

<input type="checkbox"/>	Key name	Date created
<input type="checkbox"/>	Service credentials-1	16 JUN 2020 - 06:15:13 PM
<input type="checkbox"/>	wdp-writer	16 JUN 2020 - 07:10:14 PM
<input checked="" type="checkbox"/>	Service credentials-2	17 JUN 2020 - 11:21:06 AM

At the top right of the credentials table is a 'New credential' button. On the far right, there is a vertical 'FEEDBACK' button.

- Created Jupyter Notebook in IBM Watson and import Data:

The screenshot shows the IBM Watson Studio interface. The top navigation bar includes 'Service Details', 'IBM Watson Studio', and 'Node-RED'. The main area is a Jupyter Notebook titled 'Life Expectancy Prediction'. The code in the notebook is as follows:

```
In [1]: import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from scipy.stats.mstats import winsorize
from sklearn.model_selection import train_test_split
from sklearn.linear_model import LinearRegression
from sklearn.metrics import r2_score, mean_squared_error, mean_absolute_error

In [3]: import types
import pandas as pd
from boto3.client import Config
import ibm_boto3

def __iter__(self): return 0

# @hidden_cell
# The following code accesses a file in your IBM Cloud Object Storage. It includes your credentials.
# You might want to remove those credentials before you share the notebook.
client_eab84c2c8384710bf4d452a3a8c56ce = ibm_boto3.client(service_name='s3',
    ibm_api_key_id='5QJ7L6ICHehgbMzb1UN7p0Rh06L2SgJWneecP-Zw1qho',
    ibm_auth_endpoint='https://iam.cloud.ibm.com/oidc/token',
    config=Config(signature_version='oauth'),
    endpoint_url='https://s3.eu-gb.objectstorage.service.networklayer.com')
```

The right sidebar shows the 'Data' panel with a list of files: 'Life_Expectancy.csv', 'Salary_Data.csv', 'mobile.zip', and 'watch.zip'. Each file has an 'Insert to code' button.

- Built an ML model and Created Endpoints:

The screenshot shows the IBM Watson Studio interface. The top navigation bar includes 'Service Details', 'IBM Watson Studio', and 'Node-RED'. The main area is a Jupyter Notebook titled 'Life Expectancy Prediction'. The code in the notebook is as follows:

```
Out[22]: 12.833904554965608

In [23]: mean_absolute_error(predictions, Y_test)

Out[23]: 2.5435169215233024

In [24]: from watson_machine_learning_client import WatsonMachineLearningAPIClient

2020-06-20 07:13:45,651 - watson_machine_learning_client.metanames - WARNING - 'AUTHOR_EMAIL' meta prop is deprecated. It will be ignored.

In [25]: wml_credentials = {
    "apikey": "Qwg7NXInJ682n-DEu4HQcSdiQr14ThwEI0KjBjRKhJDN",
    "iam_apikey_description": "Auto-generated for key 45b803bb-9e5a-4c34-ba85-b8bc4b84af4d",
    "iam_apikey_name": "Service credentials-3",
    "iam_role_crn": "crn:v1:bluemix:public:iam::::serviceRole:Writer",
    "iam_serviceid_crn": "crn:v1:bluemix:public:iam-identity::a/047d7e5db3284f559d41b37e6d5ea6ae::serviceid:ServiceId-bd72675-728c-4069-a2e2-98625bc71c9a",
    "instance_id": "516ce7de-7708-43d2-8e8a-c088cdf276e0",
    "url": "https://eu-gb.ml.cloud.ibm.com"
}

In [26]: client = WatsonMachineLearningAPIClient(wml_credentials)

In [27]: model_props = {
    client.repository.ModelMetaNames.AUTHOR_NAME : "Amey",
    client.repository.ModelMetaNames.AUTHOR_EMAIL : 'ameys2907@gmail.com',
    client.repository.ModelMetaNames.NAME : "Life_Expectancy"
}

In [28]: model_artifact = client.repository.store_model(model, meta_props = model_props)
```

Service Details x IBM Watson S x Node-RED : n x Node-RED D x Life Expectanc x Untitled x Student Dash x IISPS_INT_282 x + -

eu-gb.dataplatform.cloud.ibm.com/ml/models/badc1f5e-3eb9-4dac-b6e7-7fd75d4b4cb9?context=wdp&projectid=e50ff90d-2bd6-4fe9-8fe1-4a59560de54a&...

IBM Watson Studio Upgrade Amey Sawant's Account AS

My Projects / AmeyProject / Life_Expectancy

Model

Life_Expectancy

Overview Evaluation Deployments Lineage

Summary

Machine learning service	Machine Learning-pg
Model Type	scikit-learn-0.20
Runtime environment	python-3.6
Training date	20 Jun 2020, 12:43 PM
Latest version	bd79bd40-805f-492a-9667-8f98a68a96e3

Input Schema

Type here to search

1:05 PM 20/06/2020

- node-Red UI for ML:

Service Details - IBM Cloud x IBM Watson Studio x Node-RED : node-red-myqq x Node-RED Dashboard x Life Expectancy Prediction - IB x + -

node-red-myqq.eu-gb.mybluemix.net/ui/#/0?socketid=5qdrWpcOE6f1EIMNAAAB

Life Expectancy

Default

Life Expectancy 52.89573212608752

Year * 2004

Adult Mortality * 293

Infant Deaths * 87

Alcohol * 0.02

Percentage Expenditure * 15.29

Hepatitis B * 67

Measles * 466

BMI * 13.8

Under-five deaths * 120

Polio * 5

Type here to search

12:58 PM 20/06/2020

- Imported Dataset and Created AutoAI experiment:

My projects / AmeyProject / Life Expectancy Prediction

Add data source

Drop a .csv file here or [browse](#) for a file to upload.

— OR —

[Select from project](#)

Data source	Size	Columns
Life_Expectancy.csv	0.22 MB	18

Configure details

What do you want to predict?

Prediction column ①

Life expectancy x v

Prediction column: Life expectancy

PREDICTION TYPE OPTIMIZED METRIC

Regression ① **RMSE** ①

Experiment settings Run experiment

My projects / AmeyProject / Life Expectancy Prediction

Experiment summary Pipeline comparison Rank by: Root mean squared err... Score: Cross validation Holdout

Progress map ①

Prediction column: Life expectancy

```

graph LR
    A[Read dataset] --> B[Split holdout data]
    B --> C[Read training data]
    C --> D[Preprocessing]
    D --> E[Model selection]
    E --> F[Extra Trees Regressor]
    F --> G[Hyperparameter optimization]
    G --> H[Feature engineering]
    H --> I[Hyperparameter optimization]
    I --> J[P4]
    F --> K[Decision Tree Regressor]
    K --> L[Hyperparameter optimization]
    L --> M[Feature engineering]
    M --> N[Hyperparameter optimization]
    N --> O[P8]
    
```

Relationship map

Swap view

Feature engineering

DECISION TREE REGRESSOR

Started feature engineering for pipeline P7

Time elapsed: 2 minutes

[View full log](#)

Student Dashboard | IBM Watson Studio | Node-RED: node-red-myqq.eu | +

eu-gb.dataplatform.cloud.ibm.com/ml/auto-ml/acb2ea5a-1520-4ff8-9fd3-59e8d1115554/train?projectId=e50ff90d-2bd6-4fe9-8fe1-4a59560de54a&mlInstance=...

IBM Watson Studio | Upgrade | Amey Sawant's Account | AS

My projects / AmeyProject / Life Expectancy Prediction

Experiment summary | Pipeline comparison

Score: Cross validation | Holdout

Build time

Rank	Name	Algorithm	Score	Build time
1	Pipeline 3	Ext		00:00:45
2	Pipeline 4	Ext		00:00:26
3	Pipeline 1	Ext		00:00:01
4	Pipeline 2	Ext		00:00:10
5	Pipeline 7	Dec		00:00:37
6	Pipeline 8	Dec		00:00:07
7	Pipeline 5	Decision Tree Regressor	2.880	00:00:01
8	Pipeline 6	Decision Tree Regressor	2.880	00:00:01

Save as model

Save this model as a project asset so you can deploy, train, and test it.

Model name
Life Expectancy Prediction - P3 ExtraTreesRegressorEstimator

Description (optional)
Description of model

Associated project
AmeyProject

Cancel | Save

Type here to search | 10:32 AM 17/06/2020

● Built node-RED flow to integrate AutoAI:

Student Dashboard | IISPS_INT_2822_Pi | Service Details - IISPS | Service Details - IISPS | IBM Watson Studio | Node-RED: node-red-myqq.eu | Node-RED Dashboard | +

node-red-myqq.eu-gb.mybluemix.net/red/#flow/f80c9172.13372

Node-RED | Deploy

filter nodes

common

- inject
- debug
- complete
- catch
- status
- link in
- link out
- comment

function

- function
- switch
- change

Flow 1 | Flow 5 | Flow 1 | Flow 2 | Flow 1

timestamp

form

Pre Prediction

PreToken

http request

http request

Get Prediction

status

Life Expectancy

debug

6/17/2020, 12:15:53 PM node: status
msg.payload: number
57.02999992370606

Type here to search | 12:31 PM 17/06/2020