PROJECT REPORT

on

Programming Language Classification with IBM Watson Studio, Watson, and GitHub

---

Submitted by:

Ruhi Anwar Zaman (ET16BTHCS037)

Sankalpa Jyoti Bora(ET16BTHEC028)

Ankit Borgohain (ET16BTHCS050)

Gunjan Deka (ET16BTHCS045)

School of Engineering and Technology

The Assam Kaziranga University

SCHOOL OF ENGINEERING AND TECHNOLOGY

THE ASSAM KAZIRANGA UNIVERSITY

JORHAT-785006 :: ASSAM :: INDIA

CERTIFICATE

This is to certify that the project report entitled "Programming Language Classification with IBM Watson Studio, Watson, and GitHub", submitted to the School of Engineering & Technology (SET), THE ASSAM KAZIRANGA UNIVERSITY, JORHAT, ASSAM, in partial fulfilment for the completion of Semester – 7 th of the degree of Bachelor of Technology in the department of "Computer Science and Engineering" And "Electronics & Communication Engineering", is a record of bona fide work carried out by -

Miss Ruhi Anwar Zaman, Roll No. ET16BTHCS037

Mr. Sankalpa Jyoti Bora , Roll No. ET16BTHEC028

Mr. Ankit Borgohain, Roll No. ET16BTHCS050

Mr.Gunjan Deka, Roll No. ET16BTHCS045

under my guidance. All help received by us from various sources have been duly acknowledged. No part of this report has been submitted elsewhere for award of any other degree.

---------- ----------

External Mentor	**Internal Mentor**
Dr. Paras Kothari,	**Nilakshi Gogoi**
(IBM)	**(Asst. Prof. CSE Dept.)**

TABLE OF CONTENTS

Chapter 1

Introduction to Programming

Language Classifier

A model that predicts a code's programming language based on its text.But programmers can not detect every language, especially not less popular and older ones.

The most important modules are :-

- **Natural Language Processing**:-or NLP for short, is broadly defined as the automatic manipulation of natural language, like speech and text, by software. NLP is a field of Artificial Intelligence that gives machines the ability to read, understand and derive meaning from human languages.

  It is a branch under Deep learning (also known as deep structured learning or hierarchical learning)which is again a part of a broader family of machine learning methods based on artificial neural networks.

- **Watson Natural Language Classifier**:-At the core of natural language processing (NLP) lies text classification. Watson Natural Language Classifier (NLC) allows users to classify text into custom categories, at scale.

Chapter 2

Included Components

IBM Watson Studio :"Simplify and scale data science to predict and optimize your business outcomes"

IBM Watson® Studio is a leading data science and machine-learning offering built from the ground up for an AI-powered business. It helps enterprises simplify the process of experimentation to deployment, speed data exploration and model development and training, and scale data science operations across the lifecycle.

Use Watson for natural language processing, visual recognition and machine learning. Build and train AI models, and prepare and analyze data, all in one integrated environment. Intelligent data and analytic asset discovery, cataloging and governance to fuel AI apps. Build and deploy chatbots and virtual assistants.

Advantages:

1. Processes unstructured data.
2. Improve and transform customer service.
3. Handle enormous quantities of data.

Jupyter Notebook:

The Jupyter Notebook is an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and narrative text. Uses include: data cleaning and transformation, numerical simulation, statistical modeling, data visualization, machine learning, and much more.

Advantages:

- Jupyter supports over 40 programming languages, including Python, R, Julia, and Scala.
- Notebooks can be shared with others using email, Dropbox, GitHub and the Jupyter Notebook Viewer.
- Your code can produce rich, interactive output: HTML, images, videos, LaTeX, and custom MIME types.
- It is based on a set of open standards for interactive computing.
- Deploy the Jupyter Notebook to thousands of users in your organization on centralized infrastructure on- or off-site.

Chapter 3

Data Sets

The data used was generated using tools/getdata.ipynb.

Chapter 4

Steps:

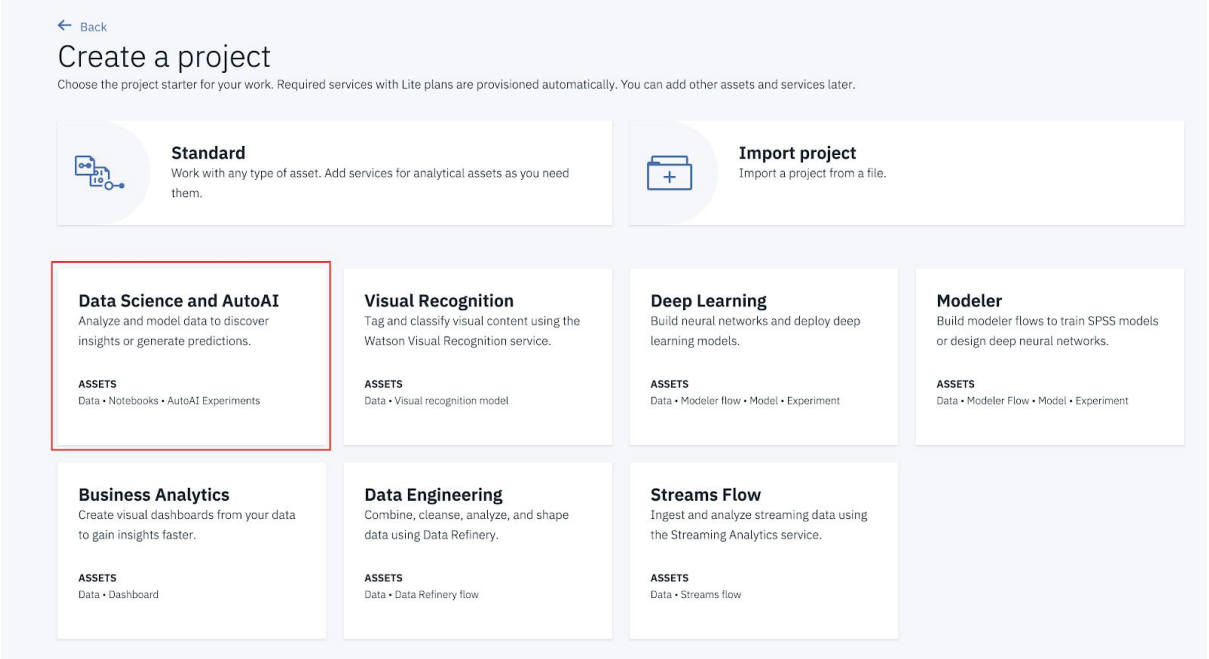
1. Create IBM Cloud services

Create the following service:

- **Natural Language Classifier**

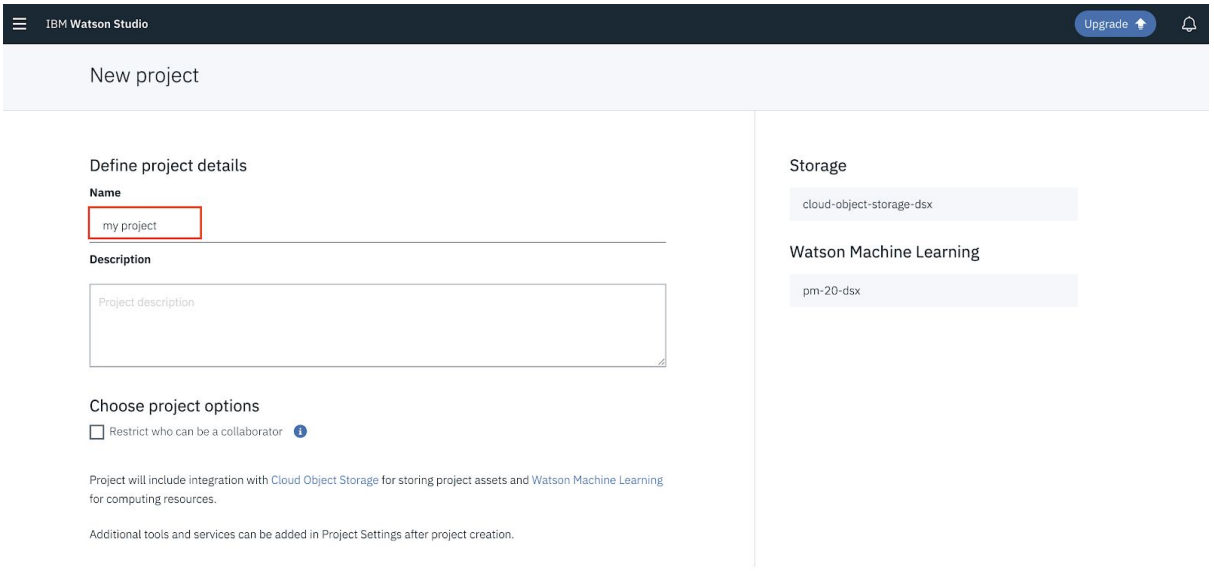
2. Create a project and add services

- **Log into IBM's Watson Studio. Once in, you'll land on the dashboard.**
- **Create a new project by clicking + New project and choosing Data Science:**

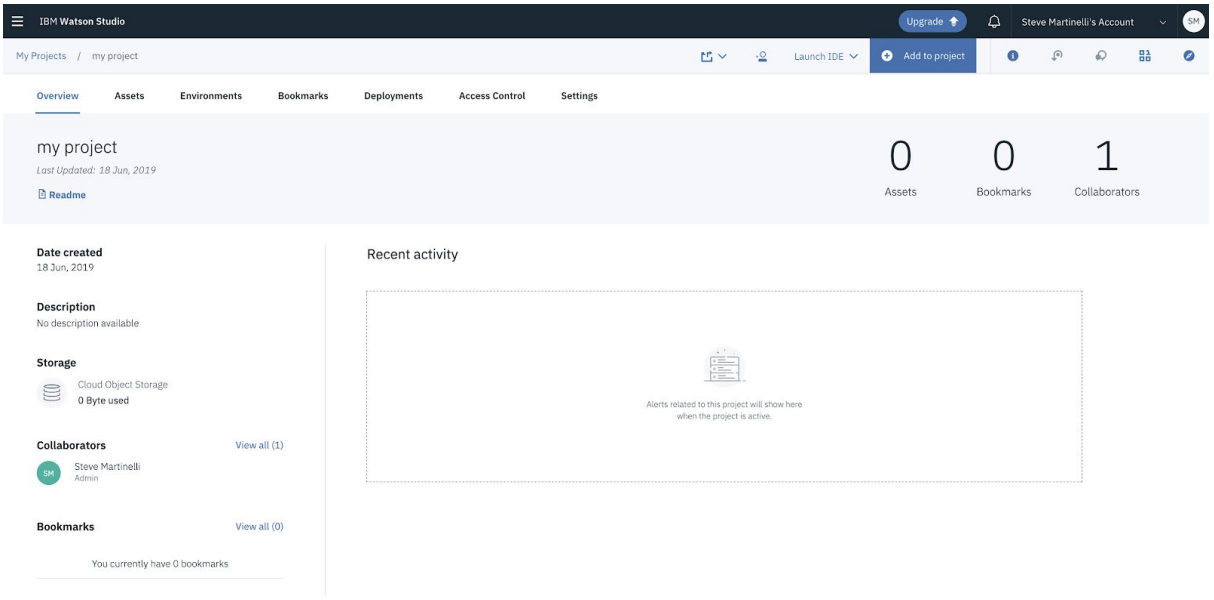


- **Enter a name for the project name and click create**
- **NOTE: By creating a project in Watson Studio a free tier Object Storage service and Watson Machine Learning service will be created in your**

IBM Cloud account. Select the free storage type to avoid fees.



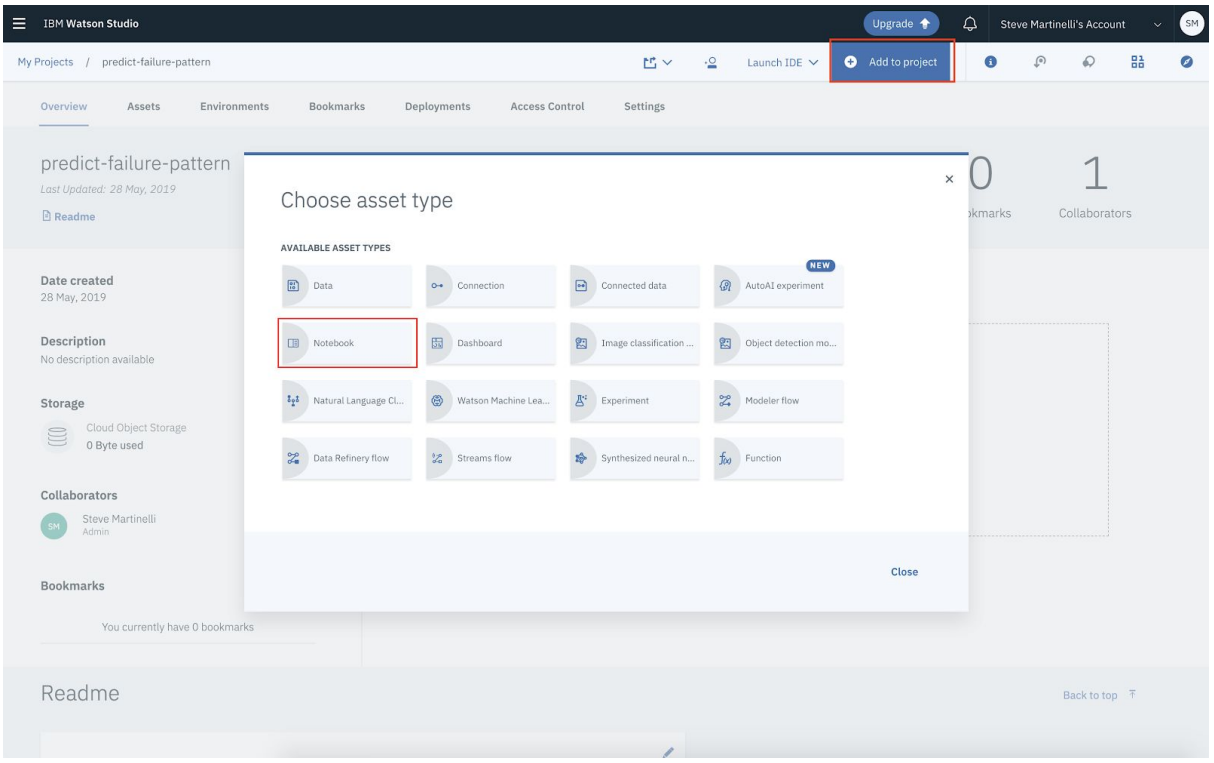
- **Upon a successful project creation, you are taken to a dashboard view of your project. Take note of the Assets and settings tabs, we'll be using them to associate our project with any external assets (datasets and notebooks) and any IBM cloud services.**



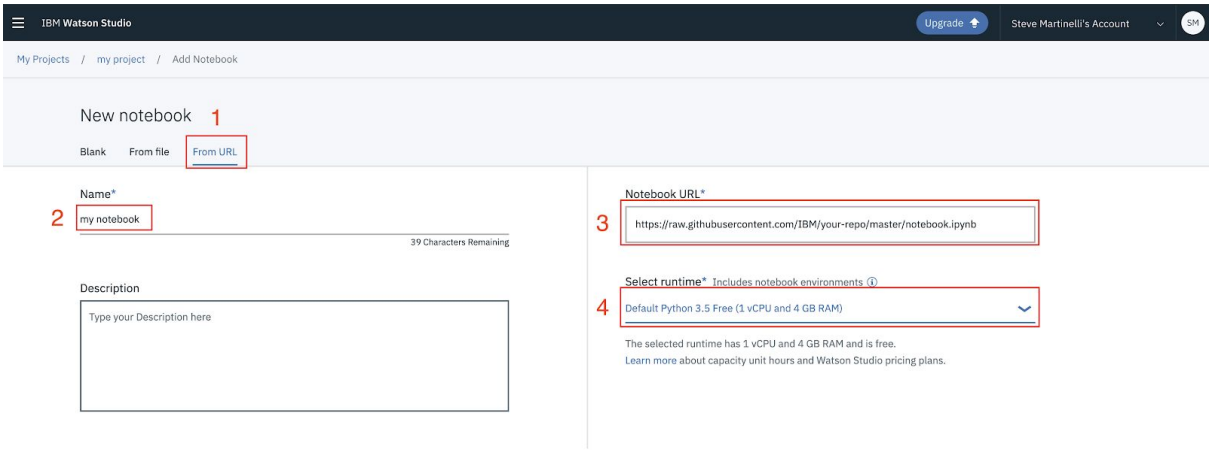
- **Associate the project with the previously created Natural Language Classifier service. Go to Settings tab in the new Project and scroll down to Associated Services. Click + and select Watson from the drop-down menu. Select an existing Watson Natural Language Classifier service or create a new one for free.**
- **Once your Natural Language Classifier(NLC) service is created, copy the credentials and save them for later, when you will use them in your Jupyter notebook.**

3. Create a notebook in Watson Studio

- **From the new project Overview panel, click + Add to project on the top right and choose the Notebook asset type.**



- **Fill in the following information:**
  - **Select the From URL tab.**
  - **Enter a Name for the notebook and optionally a description.**
  - **Under Notebook URL provide the following url: https://raw.githubusercontent.com/IBM/programming-language-classifier/master/notebooks/buildmodels.ipynb**
  - **For Runtime select the Python 3.5 option.**



-

- **Click the Create button.**
- **TIP: Once successfully imported, the notebook should appear in the Notebooks section of the Assets tab.**

4. Run the notebook in Watson Studio

When a notebook is executed, what is actually happening is that each code cell in the notebook is executed, in order, from top to bottom.

Each code cell is selectable and is preceded by a tag in the left margin. The tag format is In [x]:. Depending on the state of the notebook, the `x` can be:

- **A blank, this indicates that the cell has never been executed.**
- **A number, this number represents the relative order this code step was executed.**
- **A `*`, this indicates that the cell is currently executing.**
- **Click the `(►) Run` button to start stepping through the notebook.**
- **Continue running each cell until you finish the entire notebook**

Sample output

Chapter 5

Future AspectsNatural language classifier has a strong place in the future of computer science and human.

Following points shows the benefits and future aspects of this project

- **Automatic Summarization**: Produce a readable summary of the part of the text.(newspaper can be an example)
- Coreference Resolution: Give a sentence or larger chunk of text, determine which word refer to the same object.
- **Discourse Analysis**: This includes a number of related tasks. One task is identifying the discourse structure of the connected text.(discourse relationship between sentence)
- AI based: It is equivalent to solving the central artificial intelligence problem and making computers as intelligent as people.

Chapter 6

Conclusion

Watson Natural Language Classifier brings a sophisticated machine-learning classifier to IBM Cloud. Its approachable and intuitive REST interface makes it very easy for developers of all backgrounds to quickly train, test, and apply new classifiers to real-world problems.

Bibliography

1. https://github.com/IBM/programming-language-classifier
2. https://codeforwin.org/2017/05/programming-languages-classification.html
3. https://dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/nlc-overview.html