

Remote Summer Internship Program 2020
Machine Learning, Career Basic Program
Smartinternz

“Predicting Life Expectancy using Machine Learning”

Submitted By

By

Kunal Kokare

kunalkokare789@gmail.com

Node Red GUI Link

<https://node-red-eeiys.eu-gb.mybluemix.net/ui/>

GitHub Link

<https://bit.ly/2Y4zcnB>

Project Demonstration Video

<https://youtu.be/ye0Ev9oPrHo>

15/5/2020 - 15/6/2020

- 1. Introduction**
 - 1.1 Overview
 - 1.2 Purpose
- 2. Literature survey**
 - 2.1 Existing Problem
 - 2.2 Proposed Solution
- 3. Theoretical Analysis**
 - 3.1 Block diagram
 - 3.2 Hardware/Software Design
- 4. Experimental Investigations**
- 5. Flowchart**
- 6. Result**
- 7. Advantages & Disadvantages**
- 8. Applications**
- 9. Conclusion**
- 10. Future Scope**
- 11. Bibliography**

1. Introduction

Life expectancy plays an important role when decisions about the final phase of life need to be made. A prediction of Life Expectancy helps to analyze the average life span and thus constitute in making life decisions for the generations to come easier.

1.1 Overview

Life expectancy is a statistical measure of the average time a human being is expected to live. A typical Regression Machine Learning project 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. This problem statement provides a way to predict average life expectancy of people living in a country when various factors such as year, GDP, education, alcohol intake of people in the country, expenditure on healthcare system and some specific disease related deaths that happened in the country are given in a dataset.

The dataset used for the prediction contains data from year 2000 to 2015. It contains more than 2500 entries and around 22 columns with various features like Population, Status, Alcohol, Infant Deaths etc., which aids the prediction of the model.

1.2 Purpose

Life expectancy is one of the most important factors in end-of-life decision making. Good prognostication helps to determine the course of treatment and helps to anticipate the procurement of health care services and facilities, or more broadly: facilitates Advance Care Planning in a country. Advance Care Planning improves the quality of the final phase of life by stimulating doctors to explore the preferences for end-of-life care with their patients, and people close to the patients. Physicians, however, tend to overestimate life expectancy, and miss the window of opportunity to initiate Advance Care Planning.

2.Literature Survey

2.1 Existing Problem

Few works have been done to provide an individually customized life expectancy prediction. We have reviewed existing works and techniques in the prediction of human LE, and reached a conclusion that it is feasible to predict a PLE for individuals using evolving technologies and devices such as big data, AI, machine learning techniques, and PHDs, wearables and mobile health monitoring devices. We also identified that the collection of data will be a huge challenge due to the privacy and government policy considerations, which will require collaboration of various bodies in the health industry. The interworking of a heterogeneous health network is also a challenge for data collection. Despite these challenges, a possibility of a PLE prediction by proposing an approach of data collection and application by smartphone, with which users can enter their information to access the cloud server to obtain their own PLE, was shown.

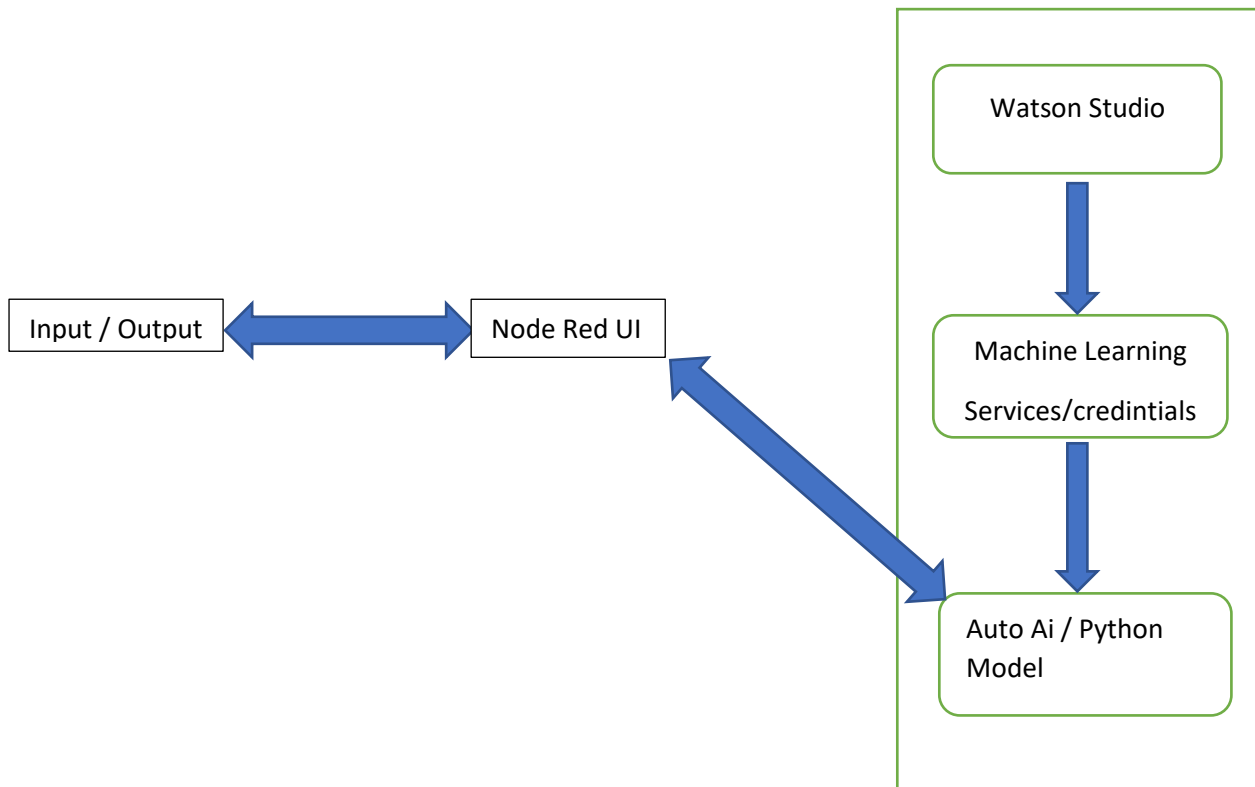
To verify the accuracy of PLE prediction and validation of data quality, big data techniques and analysis algorithms need to be developed and tested in a real-life situation with several sample groups. As artificial intelligence technology is evolving and being applied rapidly, feasibility may be increasing to collect health data from the public as well as existing health agencies such as centralized health servers.

2.2 Proposed Solution

Life expectancy is a statistical measure of the average time a human being is expected to live, Life expectancy depends on various factors: Regional variations, Economic Circumstances, Sex Differences, Mental Illnesses, Physical Illnesses, Education, Year of their birth and other demographic factors. The machine learning model built using historical data provides a way to predict average life expectancy of people living in a country when various factors such as year, GDP, education, alcohol intake of people in the country, expenditure on healthcare system and some specific disease related deaths that happened in the country are given.

3.Theoretical Analysis

3.1 Block diagram



3.2 Software Design

The regression model built in python and Auto AI model is deployed on IBM cloud. The Node-RED application then sends HTTP request with all the required parameters to the trained model. The model then sends the HTTP response which is then parsed and displayed on the UI.

4. Experimental Investigations

4.1 Following factors are taken into account for predicting the life expectancy of a country.

- Country.
- Country code.
- Year.
- Region.
- Adult mortality: Adult Mortality Rates of both sexes (probability of dying between 15 and 60 years per 1000 population).
- Infant deaths: Number of Infant Deaths per 1000 population.
- Alcohol: Alcohol, recorded per capita (15+) consumption.
- BMI: Average Body Mass Index of entire population.
- Income composition of resources: Human Development Index in terms of income composition of resources .
- Schooling: Number of years of schooling.
- Income group: Income group divided into 3 part lower , moderate and higher .

4.2 Screenshots

4.2.1 Dataset

IBM Watson Studio

Upgrade

kunal kokare's Account

My Projects / SmartInternz / life_expectancy.csv

Preview

Lineage

Schema: 12 Columns


Preview: First 1000 rows

Last refresh: 24 seconds ago

Refine

country String	country_c... String	region String	income_gr... String	year String	adult_morta... String	infant dea... String	alcohol String	Income_composition_of_resour... String	bmi String	schooling String	life_expecta... String
Afghanistan	AFG	South Asia	Low income	2000	263	62	0.01	0.479	19.1	10.1	55.12587805
Angola	AGO	Sub-Saharan Afric	Lower middle inco	2000	271	64	0.01	0.476	18.6	10	45.20478049
Albania	ALB	Europe & Centra	Upper middle inco	2000	268	66	0.01	0.47	18.1	9.9	74.27153659
United Arab Emire	ARE	Middle East & Nor	High income	2000	272	69	0.01	0.463	17.6	9.8	74.48153659
Argentina	ARG	Latin America & C	Upper middle inco	2000	275	71	0.01	0.454	17.2	9.5	73.75580488
Armenia	ARM	Europe & Centra	Lower middle inco	2000	279	74	0.01	0.448	16.7	9.2	71.27658537
Antigua and Barb	ATG	Latin America & C	High income	2000	281	77	0.01	0.434	16.2	8.9	73.39729268
Australia	AUS	East Asia & Pacific	High income	2000	287	80	0.03	0.433	15.7	8.7	79.23414634
Austria	AUT	Europe & Centra	High income	2000	295	82	0.02	0.415	15.2	8.4	78.12682927
Azerbaijan	AZE	Europe & Centra	Upper middle inco	2000	295	84	0.03	0.405	14.7	8.1	66.7582439
Burundi	BDI	Sub-Saharan Afric	Low income	2000	291	85	0.02	0.396	14.2	7.9	51.48548781
Belgium	BEL	Europe & Centra	High income	2000	293	87	0.02	0.381	13.8	6.8	77.72195122
Benin	BEN	Sub-Saharan Afric	Low income	2000	295	87	0.01	0.373	13.4	6.5	55.1882439
Burkina Faso	BFA	Sub-Saharan Afric	Low income	2000	3	88	0.01	0.341	13	6.2	50.34592683
Bangladesh	BGD	South Asia	Lower middle inco	2000	316	88	0.01	0.34	12.6	5.9	65.34797561

4.2.2 Model Using auto ai

 IBM Watson Studio

[My Projects](#) / [SmartIntenz](#) / [life expectancy - P7 ExtraTreesR...](#)

Model

life expectancy - P7 ExtraTreesRegressorEstimator

[Overview](#) [Evaluation](#) [Deployments](#) [Lineage](#)

Summary

Machine learning service	Machine Learning-26
Model Type	wml-hybrid_0.1
Runtime environment	/v4/runtimes/hybrid_0.1
Training date	3 Jun 2020, 10:18 PM

Input Schema



Column	Type
Income_composition_of_resources	"float64"
adult_mortality	"float64"
alcohol	"float64"



4.2.3 Service Credentials

Service credentials

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.

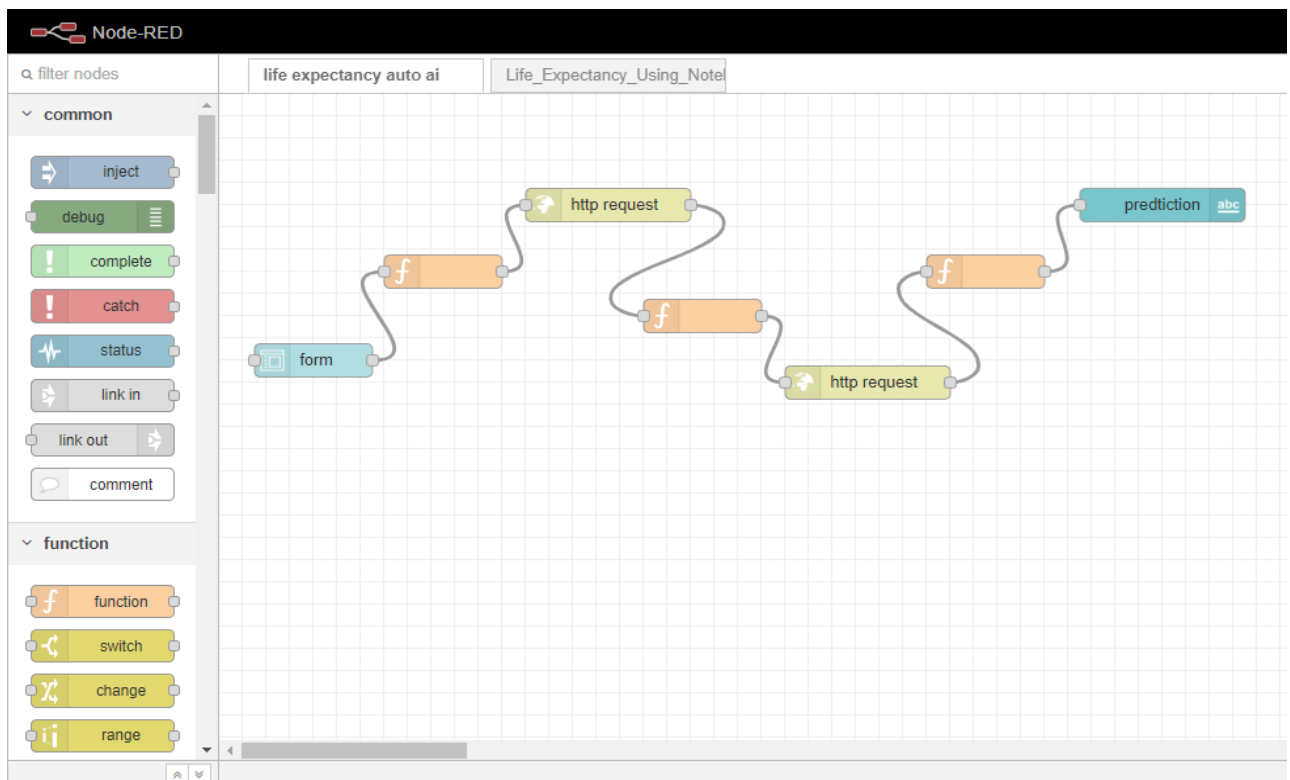
[Learn more](#)

 [New credential](#) 

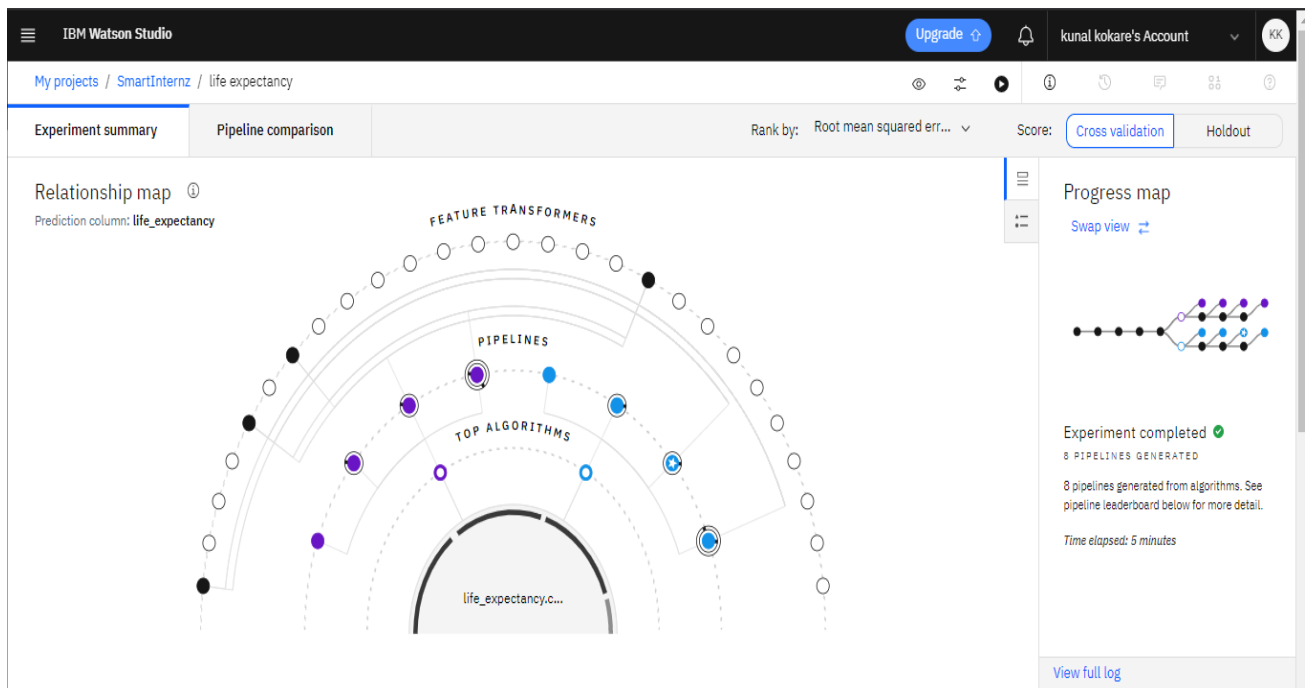
Key name	Date created	
wdp-writer	JUN 3, 2020 - 09:28:20 PM	 

```
{
  "apikey": "5RPC0kRJQDpHbfcicaYa94UdGBvxehd88cB-2-74EVC1",
  "iam_apikey_description": "Auto-generated for key db226fd3-b408-45ad-a90b-660e0e9e3ca5",
  "iam_apikey_name": "wdp-writer",
  "iam_role_crn": "crn:v1:bluemix:public:iam::::serviceRole:Writer",
  "iam_serviceid_crn": "crn:v1:bluemix:public:iam-identity::a/30a918781c634f4693f1a5fd52fc8c02::serviceid:ServiceId-a65ce64c-92d0-4772-b97e-b6b900ab4a27",
  "instance_id": "65034068-3097-4b7c-88bd-77f7cf00939e",
  "url": "https://eu-gb.ml.cloud.ibm.com"
}
```

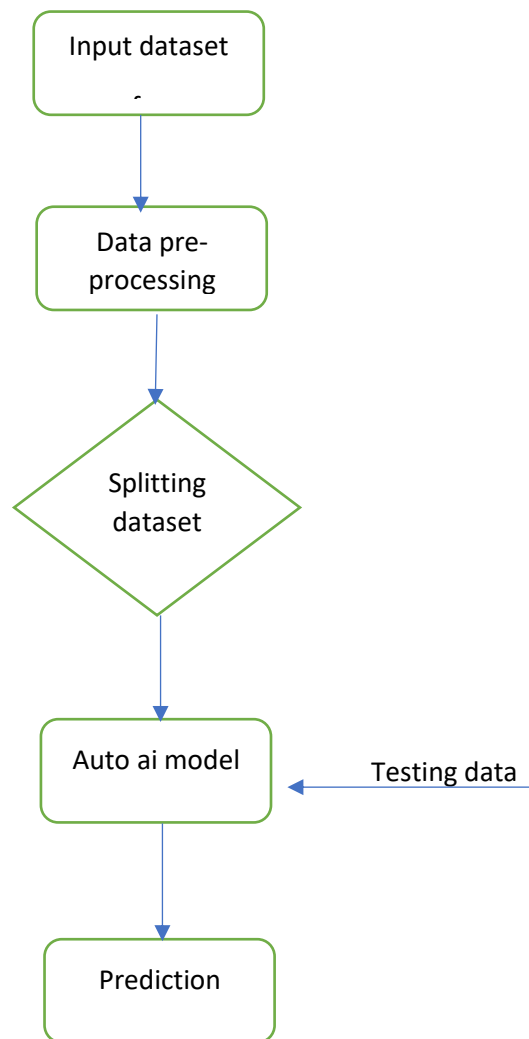

4.2.4 Node red flow



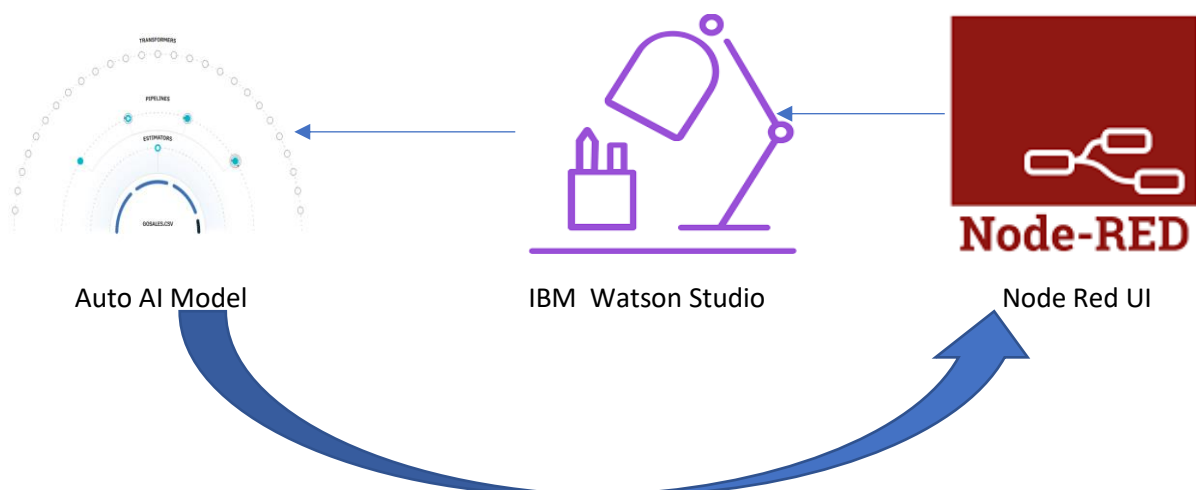
4.2.5 Feature Transformation



5.Flowchart



- Connection Between ML Model and Node red



6. Result

Life Expectancy using autoai

prediction	78.0455078125
country	India
country_code	IND
region	South Asia
income_group	higher
year	2028
adult_mortality	456
infant_deaths	21
alcohol	2
Income_composition_of_resourc...	0.45
bmi	21
schooling	45

country	India
country_code	IND
region	South Asia
income_group	higher
year	2028
adult_mortality	456
infant_deaths	21
alcohol	2
Income_composition_of_resourc...	0.45
bmi	21
schooling	15

SUBMIT**RESET**

Web based UI was developed by integrating all the services using Node red.

URL for UI Dashboard: <https://node-red-eeiys.eu-gb.mybluemix.net/ui/>

After giving the inputs for the country Afghanistan in the year 2017, the life expectancy value 78 years has been predicted.

7. Advantages & Disadvantages

Advantages:

1. Advantages of using IBM Watson:
 - Processes unstructured data
 - Fills human limitations
 - Acts as a decision support system, doesn't replace humans
 - Improves performance + abilities by giving best available data
 - Improve and transform customer service
 - Handle enormous quantities of data.
2. Sustainable Competitive Advantage
3. Easy for user to interact with the model via the UI.
4. User-friendly.
5. Easy to build and deploy.
6. Doesn't require much storage space.

Disadvantages:

- Can be only used by the people having the knowledge of data analysis.
- As the model is deployed on cloud, so one requires good internet connection to use the application.
- The Node-Red application needs to make HTTP request to IBM cloud and then another HTTP request to the model before providing the prediction. That makes the app a bit slow.

8. Applications

Life expectancy is one of the most important factors in end-of-life decision making. Good prognostication helps to determine the course of treatment and helps to anticipate the procurement of health care services and facilities, or more broadly: facilitates Advance Care Planning in a country. Advance Care Planning improves the quality of the final phase of life by stimulating doctors to explore the preferences for end-of-life care with their patients, and people close to the patients. Physicians, however, tend to overestimate life expectancy, and miss the window of opportunity to initiate Advance Care Planning.

9. Conclusion

By doing this project I have successfully created Life expectancy prediction system using IBM Watson studio, Watson machine learning and Node-RED service. Gained knowledge about IBM Cloud, IBM Watson with its services, Node-Red and how to integrate Node-Red with Machine Learning Model. The project makes a good use of machine learning in predicting life expectancy of a country. User can interact with the system via a simple user interface which is in the form of a form with input spaces which the user needs to fill the inputs into.

10. Future Scope

As future scope, we can connect the model to the database to have the record of predictions. This will help us analyze the trends in the life span.

A model with country wise bifurcation can be made, which will help to segregate the data demographically.

A mobile application can be developed that uses personal health data (from Smart Watch and Health apps) and historical data of the country that user lives in and predict the expected life span of that user.

11. Bibliography

1. Statistical Analysis on factors influencing Life Expectancy.

Source: <https://www.kaggle.com/kumarajarshi/life-expectancy-who/metadata>

2. Deploying an AutoAI model – IBM.

Source:

<https://dataplatform.cloud.ibm.com/docs/content/wsj/analyze-data/autoai-deploy-model.html>

3. Using the machine learning model – IBM.

Source: <https://cloud.ibm.com/docs/watsonknowledge-studio?topic=watson-knowledge-studio-publish-ml>

4. Welcome to the documentation of Watson Studio.

Source:

<https://dataplatform.cloud.ibm.com/docs/content/wsj/getting-started/welcomemain.html?context=analytics>

5. Infuse AI into your applications with Watson AI to make more accurate predictions

Source: <https://www.ibm.com/watson/products-services>

6. Get an understanding of Machine Learning.

Source: <https://developer.ibm.com/technologies/machine-learning/series/learning-path-machinelearning-for-developers/>