



**NextGen**  
**INVENT CORP**  
think. invent. solve.

# XGBoost Model



# Contents

<b>1. INTRODUCTION.....</b>	<b>3</b>
1.1. OVERVIEW .....	3
1.2. USE CASES .....	3
<b>2. REQUIREMENTS &amp; PREREQUISITES .....</b>	<b>4</b>
2.1. SYSTEM REQUIREMENTS.....	4
2.2. PREREQUISITES .....	4
<b>3. GETTING STARTED .....</b>	<b>5</b>
3.1. SKILL MATRIX .....	5
3.2. INSTALLATION HIERARCHY .....	5
3.3. CONFIGURATION.....	5
<b>4. LOGS .....</b>	<b>7</b>
<b>5. TROUBLESHOOTING &amp; SUPPORT .....</b>	<b>9</b>
<b>5.1 SUPPORT .....</b>	<b>9</b>
<b>APPENDIX A - RECORD OF CHANGES.....</b>	<b>10</b>
<b>APPENDIX B - ACRONYMS.....</b>	<b>10</b>
<b>APPENDIX C - REFERENCES .....</b>	<b>10</b>



### 1. Introduction

---

NextGen Invent corporation is a leading global technology services provider. Our multi-disciplinary teams, Data Scientists, Change Agents, Strategists and Process Analysis framework are delivering Robotic Process Automation solution to its clients.

NGI in partnership with Automation Anywhere is delivering Stats/ML/AI/NLP/OR/Deep Learning models as ready to use Cloud Bots that are providing efficiencies across business processes, improved customer conversion rate and reduced churn rate.

#### 1.1.Overview

A business user could use this bot to perform decision tree-based ensemble classification which is based on principles of gradient boosting framework i.e. XGBoost. The Bot takes in the input data in CSV file and performs ensemble tree methods to provide output in the MS excel.

#### 1.2.Use cases

- This bot could expedite the risk-based decisions.
- This bot could be used to classification and user-defined analytics.
- XGBoost utilizes the power of parallel processing and that is why it is much faster than GBM.



## 2. Requirements & Prerequisites

---

### 2.1. System Requirements

For the PC or server where the bot needs to run:

Device	Processor	RAM	Storage	Network
Machine	Intel Core i5 or higher	8GB or higher	16GB	1GbE
Bot Creator and Bot Runner	No additions to the machine requirements	No additions to the machine requirements	No additions to the machine requirements	No additions to the machine requirements

### 2.2. Prerequisites

**Software's required -**

Automation Anywhere Plugin version	Excel Requirement Version
11.x	2013 version or higher

**Data requirement-**

Data should be pre-processed before feeding it to the bot and it should contain at least 5,000 rows for training with around 10-12 minimum features every time to obtain desired accuracy.



## 3. Getting Started

---

### 3.1. Skill Matrix

- XGBoost is an ensemble tree method that apply the principle of boosting weak learners using the gradient descent architecture to provide powerful predictions
- This bot could be used to solve classification and user-defined analytics problems
- XGBoost has in-built regularization which prevents the model from overfitting
- XGBoost has in-built cross validation capability
- The Bot takes in the input data in CSV file and performs the XGBoost and provides output in the MS excel for quick view

### 3.2. Installation Hierarchy

- Download the bot from Bot Store.
- Double click on the .msi file.
- On Welcome to Installation wizard, click Next to continue.
- Click I agree to the terms in the license agreement radio button to accept the agreement.
- Get/Copy the License key from Bot Store Downloads into License Key, click Next to continue.
- Click Install to begin the installation.
- Click Finish to complete the installation.
- To view the installation, go to 'My Tasks' folder on AAE Client to see bot files.

### 3.3. Configuration

#### Inputs:

Parameter	Type	Direction	Additional Info
Path	String	Input	Path of the CSV file For example: C:\~\ Life expentency data.csv
ColumnCount	Numeric	Input	Count of columns available in CSV file being provided with input data
InputColumnNames	String	Input	Name of columns being provided in CSV file. For example: 'Country','Year','Status','Life expectancy','Adult Mortality','infant deaths','Alcohol','percentage expenditure','Hepatitis B','Measles','BMI','under-five deaths



			','Polio','Total expenditure', 'Diphtheria','HIV/AIDS','GDP','Population', 'thinness 1-19 years','thinness 5-9 years', 'Income composition of resources','Schooling'
InputColumnDataType	String	Input	Data Type of columns being provided in CSV file in the same order as column names. For example, 'String','Numeric','String','Numeric','Numeric','Numeric','Numeric','Numeric','Numeric','Numeric','Numeric','Numeric','Numeric','Numeric','Numeric','Numeric','Numeric','Numeric','Numeric'
MissingNumberValueReplacement	String	Input	A value that can be used by Model in case CSV file has blank field/s or missing numerical value/s.
IgnoreColumns	String	Input	Names of the column that model should ignore while calculating correlation. For example: 'Status'
PredictionColumnName	String	Input	Header of the Column on which you want to predict Ex: 'Life Expentency'

**Input Sample:**

Ex: This input examples states the life expectancy of the person from all over the world

Country	Year	Status	Life expectancy	Adult Mortality	infant deaths	Alcohol	percentage expenditure	Hepatitis B	Measles	BMI	under-five deaths	Polio
Afghanistan	2015	Developing	65	263	62	0.01	71.27962362	65	1154	19.1	83	6
Afghanistan	2014	Developing	59.9	271	64	0.01	73.52358168	62	492	18.6	86	58
Afghanistan	2013	Developing	59.9	268	66	0.01	73.21924272	64	430	18.1	89	62
Afghanistan	2012	Developing	59.5	272	69	0.01	78.1842153	67	2787	17.6	93	67
Afghanistan	2011	Developing	59.2	275	71	0.01	7.097108703	68	3013	17.2	97	68
Afghanistan	2010	Developing	58.8	279	74	0.01	79.67936736	66	1989	16.7	102	66
Afghanistan	2009	Developing	58.6	281	77	0.01	56.76221682	63	2861	16.2	106	63
Afghanistan	2008	Developing	58.1	287	80	0.03	25.87392536	64	1599	15.7	110	64
Afghanistan	2007	Developing	57.5	295	82	0.02	10.91015598	63	1141	15.2	113	63
Afghanistan	2006	Developing	57.3	295	84	0.03	17.17151751	64	1990	14.7	116	58
Afghanistan	2005	Developing		291	85	0.02	1.388647732	66	1296	14.2	118	58
Afghanistan	2004	Developing		293	87	0.02	15.29606643	67	466	13.8	120	5
Afghanistan	2003	Developing		295	87	0.01	11.08905273	65	798	13.4	122	41
Afghanistan	2002	Developing		3	88	0.01	16.88735091	64	2486	13	122	36
Afghanistan	2001	Developing		316	88	0.01	10.5747282	63	8762	12.6	122	35
Afghanistan	2000	Developing		321	88	0.01	10.42496	62	6532	12.2	122	24
Albania	2015	Developing		74	0	4.6	364.9752287	99	0	58	0	99
Albania	2014	Developing		8	0	4.51	428.7490668	98	0	57.2	1	98
Albania	2013	Developing		84	0	4.76	430.8769785	99	0	56.5	1	99
Albania	2012	Developing		86	0	5.14	412.4433563	99	9	55.8	1	99



## XGBoost Model

Polio	Total expenditure	Diphtheria	HIV/AIDS	GDP	Population	thinness 1-19 years	thinness 5-9 years	Income composition of resources	Schooling
6	8.16	65	0.1	584.2592	33736494	17.2	17.3	0.479	10.2
58	8.18	62	0.1	612.6965	327582	17.5	17.5	0.476	10.2
62	8.13	64	0.1	631.745	31731688	17.7	17.7	0.47	9.5
67	8.52	67	0.1	669.959	3696958	17.9	18	0.463	9.8
68	7.87	68	0.1	63.53723	2978599	18.2	18.2	0.454	9.5
66	9.2	66	0.1	553.3289	2883167	18.4	18.4	0.448	9.2
63	9.42	63	0.1	445.8933	284331	18.6	18.7	0.434	8.5
64	8.33	64	0.1	373.3611	2729431	18.8	18.9	0.433	8.2
63	6.73	63	0.1	369.8358	26616792	19	19.1	0.415	8.4
58	7.43	58	0.1	272.5638	2589345	19.2	19.3	0.405	8.2
58	8.7	58	0.1	25.29413	257798	19.3	19.5	0.396	7.5
5	8.79	5	0.1	219.1414	24118979	19.5	19.7	0.381	6.8
41	8.82	41	0.1	198.7285	2364851	19.7	19.9	0.373	6.5
36	7.76	36	0.1	187.846	21979923	19.9	2.2	0.341	6.2
35	7.8	33	0.1	117.497	2966463	2.1	2.4	0.34	5.5
24	8.2	24	0.1	114.56	293756	2.3	2.5	0.338	5.5
99	6	99	0.1	3954.228	28873	1.2	1.3	0.762	14.2
98	5.88	98	0.1	4575.764	288914	1.2	1.3	0.761	14.2
99	5.66	99	0.1	4414.723	289592	1.3	1.4	0.759	14.2
99	5.59	99	0.1	4247.614	2941	1.3	1.4	0.752	14.2

### Outputs:

- Output: the result in the form of the MS Excel
- Provides predictions on desired parameters

Ex: The required output predicts the life expectancy of the person using the XGBoost model baes on below class:

Gold\_Class: This column is taken as the predicted target

Predicted\_Class: This tells us the predicted age of the person

Country	Year	Status	Gold Class	Predicted Class	Adult Mortality	infant deaths	Alcohol	percentage expenditure	Hepatitis B	Measles	BMI	under-five deaths
Afghanistan	2015	Developing	65	65	263	62	0.01	71.27962362	65	1154	19.1	83
Afghanistan	2014	Developing	59.9	59.9	271	64	0.01	73.52358168	62	492	18.6	86
Afghanistan	2013	Developing	59.9	59.9	268	66	0.01	73.21924272	64	430	18.1	89
Afghanistan	2012	Developing	59.5	59.5	272	69	0.01	78.1842153	67	2787	17.6	93
Afghanistan	2011	Developing	59.2	59.2	275	71	0.01	7.097108703	68	3013	17.2	97
Afghanistan	2010	Developing	58.8	58.8	279	74	0.01	79.67936736	66	1989	16.7	102
Afghanistan	2009	Developing	58.6	58.6	281	77	0.01	56.76221682	63	2861	16.2	106
Afghanistan	2008	Developing	58.1	58.1	287	80	0.03	25.87392536	64	1599	15.7	110
Afghanistan	2007	Developing	57.5	57.5	295	82	0.02	10.91015598	63	1141	15.2	113
Afghanistan	2006	Developing	57.3	57.3	295	84	0.03	17.17151751	64	1990	14.7	116
Afghanistan	2005	Developing		57.3	291	85	0.02	1.388647732	66	1296	14.2	118
Afghanistan	2004	Developing		57	293	87	0.02	15.29606643	67	466	13.8	120
Afghanistan	2003	Developing		56.7	295	87	0.01	11.08905273	65	798	13.4	122
Afghanistan	2002	Developing		56.2	3	88	0.01	16.88735091	64	2486	13	122
Afghanistan	2001	Developing		55.3	316	88	0.01	10.5747282	63	8762	12.6	122
Afghanistan	2000	Developing		54.8	321	88	0.01	10.42496	62	6532	12.2	122
Albania	2015	Developing		77.8	74	0	4.6	364.9752287	99	0	58	0
Albania	2014	Developing		77.5	8	0	4.51	428.7490668	98	0	57.2	1
Albania	2013	Developing		77.2	84	0	4.76	430.8769785	99	0	56.5	1
Albania	2012	Developing		76.9	86	0	5.14	412.4433563	99	9	55.8	1

## 4. Logs

Error Handling In case of Errors, Error Logs & Screenshots are generated within Error Folder.

Each Bot folder contains the below hierarchy.



## XGBoost Model

### 1 Error Folder

#### 1.1 Logs

1.1.1 ErrorLogsMonth-Day-Year.txt: In case of any error, this file logs error message along with timestamp

#### 1.2 Snapshots

1.2.1 Error SnapMonth-Day-Year-Hour-Minute-SecondMillisecond.png: In case of any error, this file captures screenshot of error.

Task Status of bot is set to failed in case of error.





## 5. Troubleshooting & Support

---

### 5.1 Support

---

Please Email [bot.support@nextgeninvent.com](mailto:bot.support@nextgeninvent.com) for any support related questions.



## Appendix A - Record of Changes

---

Version #	Date of Change	Author	Notes
V1.0	June 26 <sup>th</sup> 2020	Nextgen Invent Corporation	

## Appendix B - Acronyms

---

#	Acronym	Description
1.	AA	Automation Anywhere
2.	API	Application Programming Interface
3.	RAM	Random Access Memory
4.	OS	Operating System

## Appendix C - References

---

#	Topic	Reference Link
1.	Nextgen Invent Corporation	<a href="https://www.nextgeninvent.com/">https://www.nextgeninvent.com/</a>

