

SMS analysis using Watson Knowledge Studio

Introduction

SMS analysis using the Watson Knowledge Studio project examines the textual messages and extracts meta-data like keywords, concepts, categories, emotions, semantic rules, sentiments, and any domain-specific requirements. The custom model uses a machine learning technique called Text Analysis. The flexible and collaborative environment provided by the Watson Knowledge studio makes it possible to train the custom model based on automatic annotations and easily apply it to any new document.

Manually processing and organizing text data takes time, it's tedious, inaccurate, and it can be expensive if you need to hire extra staff to sort through text. Since the custom machine learning model is trained on real-life text samples, it makes it easier to collect and extract the entity types from a large textual document, thus reducing cost and essentially time.

Theoretical Analysis

Natural Language Processing

Natural language processing (NLP) refers to the branch of computer science, and more specifically, the branch of Artificial Intelligence concerned with giving computers the ability to understand the text and spoken words in much the same way human beings can.

NLP combines computational linguistics rule-based modeling of human language with statistical, machine learning, and deep learning models. Together, these technologies enable computers to process human language in the form of text or voice data and to 'understand' its full meaning, complete with the speaker or writer's intent and sentiment.

NLP drives computer programs that translate text from one language to another,

respond to spoken commands, and summarize large volumes of text rapidly even in real-time. There's a good chance you've interacted with NLP in the form of voice-operated GPS systems, digital assistants, speech-to-text dictation software, customer service chatbots, and other consumer conveniences. But NLP also plays a growing role in enterprise solutions that help streamline business operations, increase employee productivity, and simplify mission-critical business processes.

Natural Language Understanding

Natural Language Understanding is a subtopic of Natural Language Processing in artificial intelligence that deals with machine reading comprehension. The IBM Watson Natural Language Understanding service enables developers to extract insights from unstructured text to power a new generation of cognitive applications. It is used to analyze text and extract meta-data like keywords, concepts, sentiments, relations, keywords, categories, semantic rules, or any domain-specific customizations

Watson Knowledge Studio

It is an interface where we can teach the IBM Watson the language of our domain with custom models that identify entities and relationships unique to our industry.

It is used to build models in a collaborative environment designed for both developers and domain experts, without needing to write code.

We can create machine learning models that understand linguistic meaning or a rule-based model that finds entities and models based on your domain

We can train our models to identify certain entities and define custom relation types.

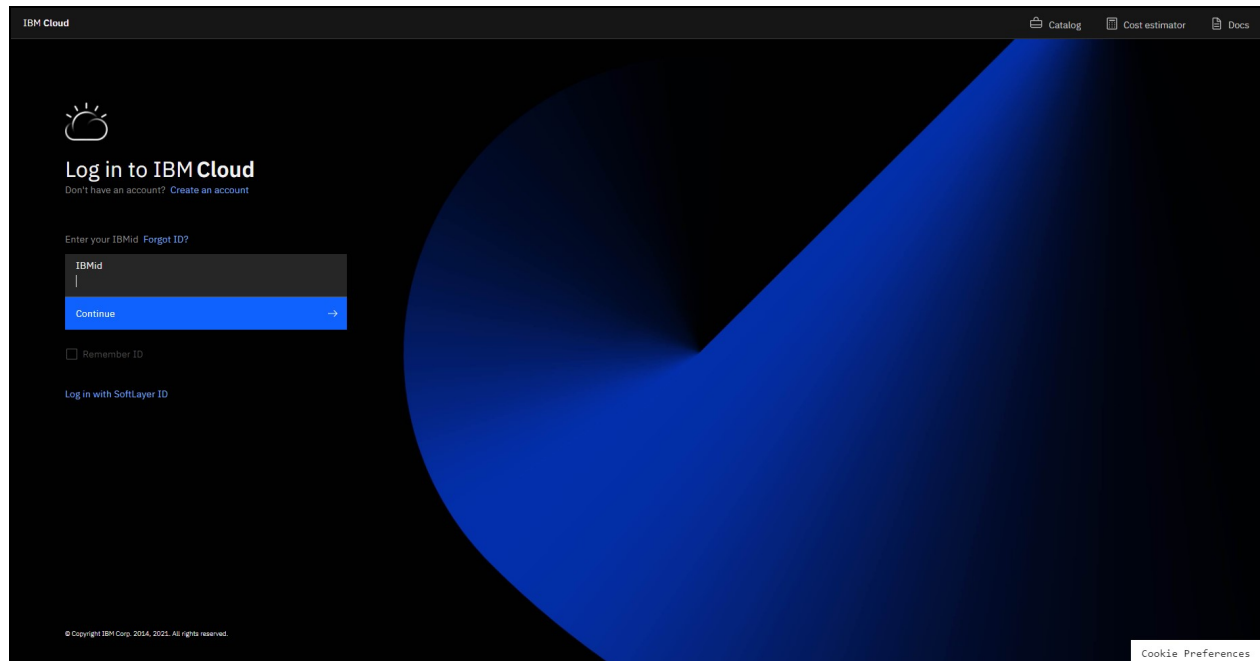
Node-RED

Node-RED is a programming tool for wiring together hardware devices, APIs, and online services in new and interesting ways. It provides a browser-based editor that makes it easy to wire together flows using the wide range of nodes in the palette that can be deployed to its runtime in a single click. JavaScript functions can be created within the editor using a rich text editor. A built-in library allows you to save useful functions, templates, or flows for re-use. The flows created in Node-RED are stored using JSON which can be easily imported and exported for sharing with others.

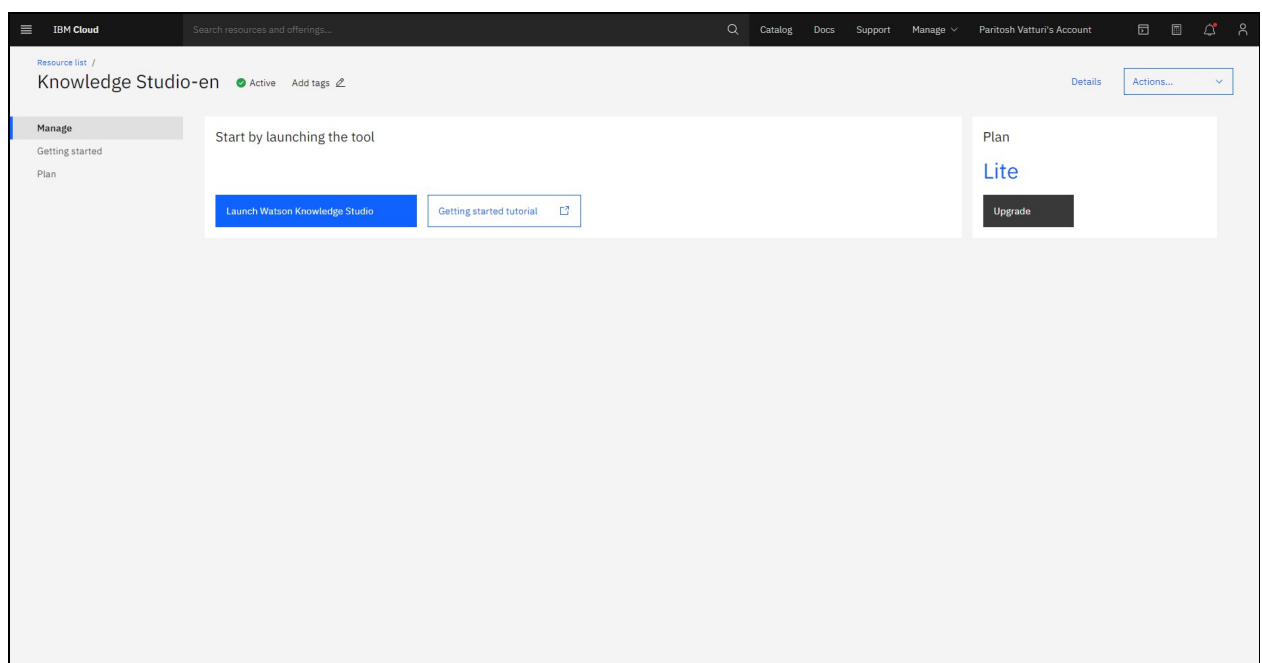
Procedure

The project flow contains the following steps -

- Create IBM account



- Create Watson Knowledge Studio Service



Create Watson Natural language Understanding Service

The screenshot shows the IBM Cloud console for the 'Natural Language Understanding-8l' service. The service is in an 'Active' state. The left sidebar contains a 'Manage' section with links for 'Getting started', 'Service credentials', 'Plan', and 'Connections'. The main content area has a 'Start by viewing the tutorial' section with links to 'Getting started tutorial' and 'API reference'. Below this is a 'Credentials' section with fields for 'API key' and 'URL'. The 'API key' field is masked with dots, and the 'URL' field contains the value 'https://api.us-south.natural-language-understanding.watson.cloud.ibm.com/instances/36a...'. To the right, there is a 'Plan' section showing the 'Lite' plan with an 'Upgrade' button.

- Load entity type system in Watson Knowledge Studio

The screenshot shows the IBM Watson Knowledge Studio console for the 'Entity Types' configuration. The left sidebar contains a 'Back to Workspaces' link and a list of assets: 'Documents', 'Entity Types', 'Relation Types', and 'Dictionaries'. The 'Entity Types' asset is selected. The main content area has a 'Entity Types' section with a '14' badge. Below this, there are 'Add Entity Type' and 'Upload' buttons. To the right, there is a 'Download Types' button. Below these buttons is a search bar with the text 'Enter text to filter'. Below the search bar is a table with the following columns: 'Entity Type Name', 'Roles', 'Subtypes', and 'Action'. The table contains 14 rows of entity types, each with a checkbox, a name, a role, a subtype, and an action (Edit/Delete). The table is paginated with 'Items per page: 10' and '1-10 of 14 items'. The bottom right of the table shows '1' of 2 pages.

<input type="checkbox"/>	Entity Type Name	Roles	Subtypes	Action
<input type="checkbox"/>	Terms_Conditions	Terms_Conditions		Edit Delete
<input type="checkbox"/>	Merchant_URL	Merchant_URL		Edit Delete
<input type="checkbox"/>	Validity_Period	Validity_Period		Edit Delete
<input type="checkbox"/>	Coupon_Code	Coupon_Code		Edit Delete
<input type="checkbox"/>	Festival_Occasion_Offers	Festival_Occasion_Offers		Edit Delete
<input type="checkbox"/>	Offer_Branch	Offer_Branch		Edit Delete
<input type="checkbox"/>	Offer_Applicable	Offer_Applicable		Edit Delete
<input type="checkbox"/>	Offer_Product	Offer_Product		Edit Delete
<input type="checkbox"/>	Points	Points		Edit Delete
<input type="checkbox"/>	Cashback	Cashback		Edit Delete

- Load training data files into Watson Knowledge Studio

IBM Watson Knowledge Studio

Documents

Document Sets (9) Documents (All, 120)

Upload Document Sets To begin annotating documents, go to [Annotations](#) page. Download Document Sets

Name	Documents	Last Modified	Action
All	120	-	
SecondSmsSet.txt_set	20	02/27/2018	Rename Delete
FirstSmsSet.txt_set	20	02/28/2018	Rename Delete
ThirdSmsSet.txt_set	20	02/27/2018	Rename Delete
FourthSmsSet.txt_set	20	02/28/2018	Rename Delete
FifthSmsSet.txt_set	40	02/28/2018	Rename Delete
Training	84	06/01/2021	Delete
Test	27	06/01/2021	Delete
Blind	9	06/01/2021	Delete

Items per page: 10 1-9 of 9 items 1 of 1 pages

- Generate model by training and evaluating data.

IBM Watson Knowledge Studio

Performance

SMS-Analysis

Language of documents: English

Number of documents per set

84 Training Set
27 Test Set
9 Blind Set

Training Set [View Ground Truth](#)
Test Set [View Ground Truth](#)
[View Decoding Results](#)

Last trained on: Jun 1, 2021 10:19:42 PM
Last evaluated on: Jun 1, 2021 10:21:43 PM

[Train and evaluate](#) Low performance? Click here to train.

Document set evaluation ⓘ

Model over time [View Log](#)

Score

Version

Mention: 0.70 Precision: 0.70 Recall: 0.70
Relation: 0.48 Precision: 0.56 Recall: 0.42
Coreference: -- Precision: -- Recall: --

Low performing range

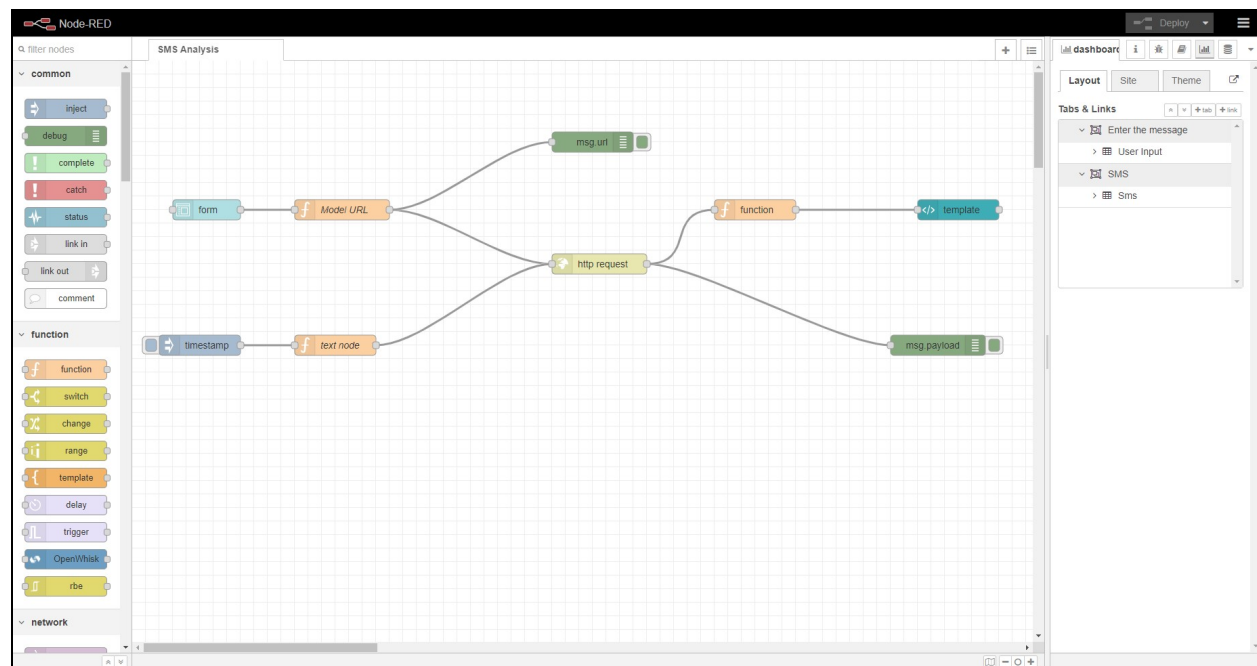
- The Watson Knowledge Studio model is deployed to Watson NLU.

The screenshot shows the 'Versions' page in IBM Watson Knowledge Studio. The left sidebar contains a navigation menu with options like 'Assets', 'Documents', 'Entity Types', 'Relation Types', 'Dictionaries', 'Rule-based Model', 'Machine Learning Model', 'Pre-annotation', 'Annotations', 'Performance', 'Versions' (selected), 'Settings', and 'Help'. The main content area is titled 'Versions' and has a sub-tab 'Machine Learning Model'. It contains two boxes: 'Machine learning models, once trained, can be used to annotate new documents prior to human annotation.' with a 'Go to Pre-annotation page' button, and 'Export the current version of your model to use in other Watson applications such as Watson Explorer.' with an 'Export current model' button. Below these is a 'Version History and Deployment' table.

Version	Base	Creation Date	Entity Scores	Relation Scores	Description	Action
1.2		Current Version	0.70 (0.70 / 0.70)	0.48 (0.56 / 0.42)		Create Version
1.1		06/01/2021	0.70 (0.70 / 0.70)	0.48 (0.56 / 0.42)		Promote Delete Deploy

Below the table, it indicates 'Deployed Models (1)'.

- Build Node-RED application (UI)



- The SMS message is analyzed by Watson NLS for processing and returns extracted domain-specific entities

SMS

Sms

Analysed SMS

Type	Text	Confidence
Merchant_URL	http://bit.ly/2cZArOK	0.999215
Offer	get Rs.75 FREE	0.998104
Coupon_Code	INDIAN	0.991606
Merchant	MobiKwik	0.991091
Offer	save Rs.100 on recharges	0.990855

text *

SUBMIT

CANCEL

Conclusion

Our team has successfully completed the project SMS analysis using Watson Knowledge Studio. By all counts and proven results, this model is efficient in extracting entity types from text messages.

We would like to thank all the mentors from SmartBridge for their support and guidance throughout the course of the project.

Thank you

