

Project Development Phase
Model Performance Test

Date	10 November 2022
Team ID	PNT2022TMID01193
Project Name	AI BASED DISCOURSE FOR BANKING INDUSTRY
Maximum Marks	10 Marks

Model Performance Testing:

Project team shall fill the following information in model performance testing template.

S.No.	Parameter	Values	Screenshot
1.	Model Summary	-	<p>There are two different tasks at the core of a chatbot:</p> <ul style="list-style-type: none">• User request analysis• Returning the response <p>User Request Analysis</p> <p>This is the first measure a chatbot will take in order to fully understand what you are trying to say. The chatbot will analyze your request and identify any keywords or phrases that can help it better understand your intentions. It can also extract any important information you mention in your message.</p> <p>The ability to identify the user's intent and extract data and relevant entities contained in the user's request is the first condition and the most important step at the core of a chatbot: if you are not able to correctly process their requests, you will not be able to answer them. you won't be able to provide the correct answer.</p> <p>Returning the Response</p> <p>Once the user's intent has been identified, the chatbot must provide the most appropriate response for the user's request. The answer may be:</p> <ul style="list-style-type: none">• A generic and predefined text;• A text retrieved from a knowledge base that contains different answers;• A contextualized piece of information based on data the user has provided;• Data stored in enterprise systems;• The result of an action that the chatbot performed by interacting with one or more backend applications; or,• A disambiguating question that helps the chatbot to correctly understand the user's request.

2.	Accuracy	<div>Training Accuracy =95.5</div> <div>Validation Accuracy = 0.8045</div>	<div><div><div><div>Watson Blog</div><div>Home</div><div>Watson Assistant</div><div>Pricing</div><div>Client stories</div><div>Get started free</div></div><div><div>the time to production. We're excited to announce that Watson Assistant has a new and improved intent detection algorithm, which is more accurate versus commercial and open-source solutions in a recently published benchmark (see Table).3</div><table><thead><tr><th></th><th>Fold average</th><th>Subtask average</th><th>Average</th></tr></thead><tbody><tr><td>IBM Watson Assistant</td><td>79.8</td><td>76.9</td><td>78.2</td></tr><tr><td>Google Dialogflow</td><td>68.2</td><td>64.0</td><td>66.6</td></tr><tr><td>Microsoft LUIS</td><td>55.4</td><td>52.8</td><td>54.2</td></tr><tr><td>Haystack</td><td>73</td><td>67.6</td><td>70.2</td></tr><tr><td>Dialog</td><td>67.4</td><td>58.8</td><td>63.2</td></tr><tr><td>BOBY</td><td>75.9</td><td>64.3</td><td>69.9</td></tr></tbody></table><div>Because of these improvements, the accuracy of Watson Assistant's latest model is 79%, up from 76.3% in the immediately previous version. This means a Watson</div></div></div><div><div>result of my test using the Car Dashboard workspace and 5 folds:</div><div><div>Accuracy: 0.8929</div><div>Accuracy: 0.8889</div><div>Accuracy: 0.9147</div><div>Accuracy: 0.9167</div><div>Accuracy: 0.9127</div><div>Average accuracy: 0.9052</div><div>Woof Well done.</div></div></div></div>		Fold average	Subtask average	Average	IBM Watson Assistant	79.8	76.9	78.2	Google Dialogflow	68.2	64.0	66.6	Microsoft LUIS	55.4	52.8	54.2	Haystack	73	67.6	70.2	Dialog	67.4	58.8	63.2	BOBY	75.9	64.3	69.9
	Fold average	Subtask average	Average																												
IBM Watson Assistant	79.8	76.9	78.2																												
Google Dialogflow	68.2	64.0	66.6																												
Microsoft LUIS	55.4	52.8	54.2																												
Haystack	73	67.6	70.2																												
Dialog	67.4	58.8	63.2																												
BOBY	75.9	64.3	69.9																												