**Performance Evaluation: Precession, Recall & F1 Score**

**Confusion Matrix**:

Where is “N” is no of prompted queries to the Model.

|  |  |  |
| --- | --- | --- |
| **N=n** | Predicted No (Response Match) | Predicted Yes  (Response Match) |
| Actual No  (Ground truth Match) | True Negative | False Positive |
| Actual Yes  (Ground Truth Match) | False Negative | True Positive |

**False Positive:** If the data is correct and we are able to understand the context, but it doesn’t fully match with the data available in the ground truth.

**True Positive:** If everything matches with the ground truth with semantic meaning.

**False Negative:** if data is correct in the ground truth and we didn’t get a response as per the expected context.

**True Negative:** It is always considered as zero because ground truth cannot be false.

1. **WHO1: Title': 'Title**

|  |  |  |
| --- | --- | --- |
| **N=46** | Predicted No | Predicted Yes |
| Actual No | 0 | 0 |
| Actual Yes | 0 | 46 |

**Precision = 46/46+0 = 1**

**Recall = 46/46 = 1**

**F1 Score = 2(1\*1/1+1) = 1**

1. **WHY1: What is the evaluation purpose in this study?**

|  |  |  |
| --- | --- | --- |
| **N=46** | Predicted No | Predicted Yes |
| Actual No | 0 | 6 |
| Actual Yes | 1 | 39 |

**Precision = 39/39+6 = 0.86**

**Recall= 39/39+1 = 0.97**

**F1 Score = 2\*(0.832/1.83) = 0.90**

1. **WHY2: What is the research aim in this study?**

|  |  |  |
| --- | --- | --- |
| **N=46** | Predicted No | Predicted Yes |
| Actual No | 0 | 38 |
| Actual Yes | 2 | 6 |

**Precision = 6/38+6= 0.13**

**Recall = 6/38+2 = 0.15**

**F1 Score = 2\*(0.0195/0.28) = 0.139**

1. **WHERE1: By which commercial Cloud provider(s) are the evaluated services supplied?**

|  |  |  |
| --- | --- | --- |
| **N=46** | Predicted No | Predicted Yes |
| Actual No | 0 | 1 |
| Actual Yes | 0 | 45 |

**Precision = 45/45+1 = 0.97**

**Recall = 45/45+0 = 1**

**F1 Score = 2(0.97/1.97) = 0.98**

1. **WHERE2: What commercial Cloud computing services were evaluated?**

|  |  |  |
| --- | --- | --- |
| **N=46** | Predicted No | Predicted Yes |
| Actual No | 0 | 42 |
| Actual Yes | 1 | 3 |

**Precision = 3/42+3 = 0.066**

**Recall = 3 / 3+1 = 0.75**

**F1 Score = 2(0.066\*0.75/)0.816 = 0.121**

1. **WHAT1: What feature(s) was/were evaluated in this study?**

|  |  |  |
| --- | --- | --- |
| **N=46** | Predicted No | Predicted Yes |
| Actual No | 0 | 27 |
| Actual Yes | 5 | 14 |

**Precision = 14/14+27 = 0.34**

**Recall = 14/14+5 = 0.73**

**F1 Score = 2(0.248/1.07) = 0.46**

1. **WHAT2: What components were concerned for the evaluated feature?**

|  |  |  |
| --- | --- | --- |
| **N=46** | Predicted No | Predicted Yes |
| Actual No | 0 | 31 |
| Actual Yes | 0 | 15 |

**Precision = 15/46 = 0.32**

**Recall = 15 /15+0 =1**

**F1 Score = 2(0.32/1.32) = 0.48**