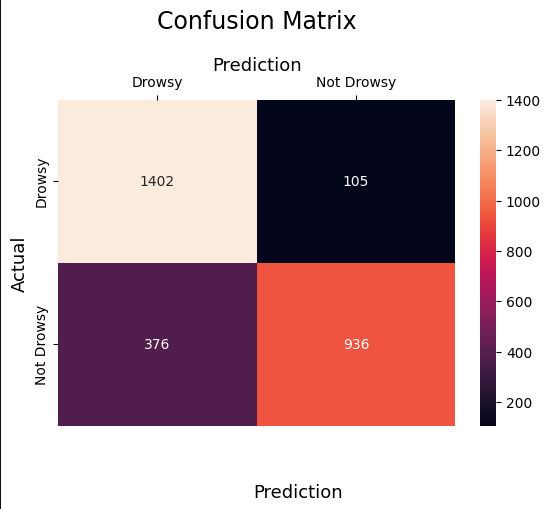
**(ii).** **Analysis on the Effectiveness of Knowledge Distillation.**

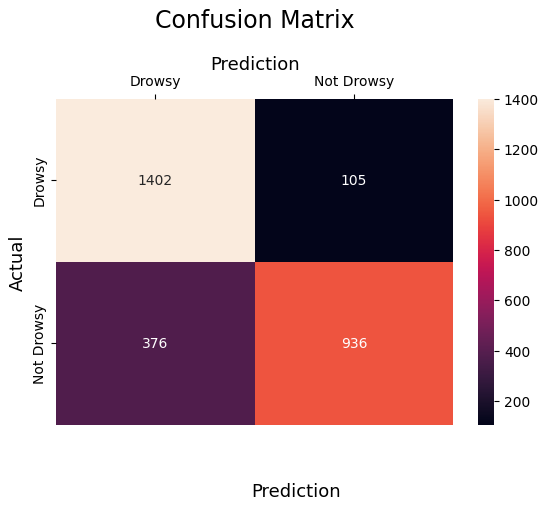
Performance of the Teacher and the Student Models

**a. Confusion Matrix:**

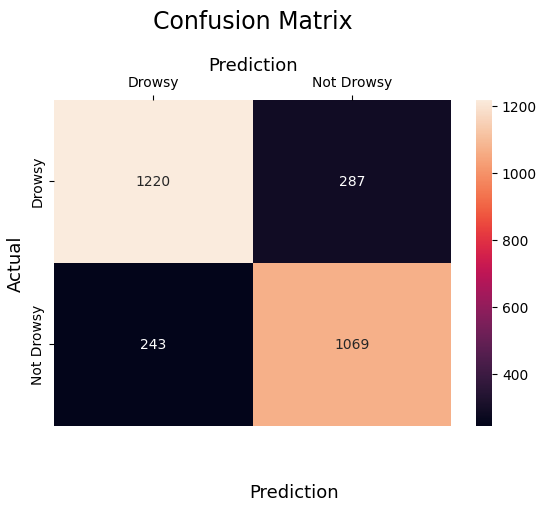
Teacher Model:



Student Model(Before Distill):



Student Model(After Distill):



**b. Tabulation of performance Metrics related to model(teacher’s performace is same as student- teacher’s performance should be superior )**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Accuracy | Precision | Recall | F1-Score |
| Proposed Capsule MobileNet  **(Teacher)** | 82.94% | 0.8400 | 0.8294 | 0.8266 |
| Lightweight version of Proposed  **(Student Before Distillation)** | 82.94% | 0.8400 | 0.8294 | 0.8266 |
| **Student after distillation** | 81.20% | 0.8127 | 0.8120 | 0.8121 |

c.  **Tabulation of performance Metrics related to suitability of edge devices – when the architecture of student is same before and after., latency won’t change ., check this result too.,**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Models | Model Size(MB) | Inference Latency(Sec) | Compression Ratio | Flops(GFLOPs) |
| Proposed Capsule MobileNet  **(Teacher)** | 43.40 | 45.5690 | - | 0.760 |
| Lightweight version of Proposed  **(Student Before Distillation)** | 20.58 | 44.3563 | 2.11 | 0.760 |
| **Student after distillation** | 20.58 | 39.9427 | 2.11 | 0.760 |

Ppl ., there can be a change in inference latency if the student model’s architecture is changed., but thts not the case with us .,

**Our claim is ., for a decent accuracy we obtain reduced FLOPS and inference latency thru distillation**

**There might be some changes obtained due to batch size changes ., but In our case I think we are not altering the batch size.,**