***Classification of Events***

**True Positives (TP) - Correct Predictions by LLM**

These are events that the LLM correctly predicted and also happened during the real Battle of Stalingrad.

| **Event** | **Definition** |
| --- | --- |
| The battle was a turning point in World War II | Stalingrad marked a major shift in momentum, placing Germany on the defensive for the rest of the war. |
| The defending forces (Soviet Union) used urban warfare tactics to their advantage | The Soviet Army fortified buildings, used snipers, and turned the city into a defensive fortress. |
| The attacking forces (Germany) suffered from extended supply lines | The Wehrmacht was stretched too thin, leading to major logistical problems. |
| The battle was characterized by brutal, close-quarters combat | Fighting in buildings, streets, and factories resulted in intense hand-to-hand and room-to-room battles. |
| The harsh winter played a decisive role in the battle’s outcome | The Soviet troops were better prepared for winter, while German forces suffered from cold and supply shortages. |
| The Soviet counteroffensive encircled the German Sixth Army | Operation Uranus encircled the German forces, cutting them off from reinforcements and supplies. |
| The German leadership refused to allow a tactical retreat | Hitler insisted on holding the city at all costs, which led to the destruction of the German Sixth Army. |
| The Luftwaffe was unable to supply the encircled German forces effectively | German air supply efforts failed, leaving trapped soldiers without enough food or ammunition. |
| Soviet reinforcements from other fronts helped secure victory | The Red Army brought in fresh reinforcements to surround and destroy the German army. |
| The battle resulted in massive casualties on both sides | Over 2 million casualties, making it one of the bloodiest battles in history. |

Total True Positives (TP): **10**

**False Positives (FP) - Incorrect Predictions by LLM**

These are events that the LLM predicted but did NOT actually happen in the real Battle of Stalingrad.

| **Event** | **Why It Did NOT Happen?** |
| --- | --- |
| The Soviet Union launched a preemptive attack to prevent the German invasion | The battle began with a German offensive; the Soviets did not strike first. |
| The battle included large-scale tank engagements in open fields | While tanks played a role, Stalingrad was mostly urban warfare, not large open-field tank battles. |
| The battle involved a major naval engagement | There was no significant naval battle; the Volga River was used for supply routes, but no fleet combat occurred. |
| The Soviets received large amounts of Western reinforcements during the battle | While Lend-Lease aid helped, direct reinforcements from the West were not a factor during the battle. |
| Germany attempted to negotiate a ceasefire to evacuate its forces | Hitler refused to allow surrender until it was too late, and no serious negotiations took place. |
| The Soviet Union relied heavily on aerial bombing campaigns to weaken German positions | The Soviets focused more on ground combat and artillery rather than strategic bombing. |
| The battle ended with a quick German retreat | The Germans were trapped and forced to surrender; there was no organized retreat. |
| The battle was influenced by major guerrilla warfare operations behind German lines | While partisan activity existed, it was not a decisive factor in the battle itself. |

Total False Positives (FP): **8**

**False Negatives (FN) - Missed Predictions by LLM**

These are real events that happened during the Battle of Stalingrad but were NOT predicted by the LLM.

| **Event** | **Definition** |
| --- | --- |
| The battle began with heavy German bombing, reducing much of the city to rubble | The Luftwaffe conducted massive air raids, which ironically helped the Soviets by creating defensive ruins. |
| Soviet women played a key role in frontline combat and logistics | Thousands of female snipers, medics, and tank crew members contributed significantly to the battle. |
| German forces faced starvation and disease as the siege progressed | With supplies cut off, German soldiers resorted to eating horses and even saw cases of cannibalism. |
| The Soviets used “hugging” tactics to neutralize German airpower | Soviet troops stayed as close as possible to German positions, preventing effective Luftwaffe bombings. |
| The Germans attempted a failed breakout (Operation Winter Storm) | A relief attempt led by Field Marshal von Manstein was launched but ultimately failed. |
| Soviet deception operations (maskirovka) played a key role in misleading German intelligence | The Soviets used fake movements and misinformation to hide their buildup for the counteroffensive. |
| Thousands of Axis prisoners died in Soviet captivity after the battle | Of the 91,000 German troops captured, fewer than 6,000 survived to return home. |
| Stalingrad became a symbol of Soviet resilience and propaganda | The victory was heavily used in Soviet propaganda to boost morale and justify further offensives. |

Total False Negatives (FN): **8**

***Final Results***

| **Metric** | **Value** |
| --- | --- |
| Total True Positives (TP) | **10** |
| Total False Positives (FP) | **8** |
| Total False Negatives (FN) | **8** |
| Precision | **55.56%** |
| Recall | **55.56%** |
| F1-Score | **55.56%** |

***Critical Analysis of LLM’s Prediction***

1. Major Issue: The LLM Predicted the Wrong Winner

* The LLM incorrectly assumed the attacking force (Germany) won, when in reality, the defending force (Soviet Union) decisively won.
* Germany lost the Sixth Army, leading to a major strategic defeat.

2. Strengths in LLM’s Predictions

Correctly identified major aspects of the battle:

* Urban warfare heavily favored defenders.
* Winter conditions played a crucial role.
* The German supply lines were overextended.
* The Soviet counteroffensive trapped the German army.

3. Weaknesses in LLM’s Predictions

Overestimated German capabilities:

* Did not account for failed German supply efforts.
* Ignored Hitler’s refusal to retreat.
* Underestimated the Soviet ability to encircle and counterattack.

Missed key moments:

* Luftwaffe bombings destroyed the city but helped Soviet defenses.
* Soviet deception operations (maskirovka) were highly effective.

Invented historical inaccuracies:

* No preemptive Soviet attack.
* No large-scale tank battles.
* No ceasefire negotiations.

***Final Verdict***

* The LLM’s prediction was weak (~55.56% F1-score) due to incorrect outcome prediction and historical inaccuracies.
* Improving recall (identifying more real-world details) and correcting historical errors would significantly improve future assessments.