***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, putting Germany on the defensive for the rest of the war. |
| The battle was primarily fought in an urban environment | Fighting occurred in streets, buildings, and factories, making urban warfare the dominant mode of combat. |
| The defending forces (Soviet Union) utilized snipers effectively | Soviet snipers, notably Vasily Zaitsev, played a major role in weakening German morale and forces. |
| The attacking forces (Germany) suffered from extended supply lines | The Wehrmacht’s logistical issues worsened as Soviet counterattacks cut off critical supplies. |
| The battle took place during a harsh winter, affecting combat operations | Soviet forces were better adapted to winter warfare, while German troops suffered from extreme cold. |
| The Soviet Union launched a successful counteroffensive | Operation Uranus encircled the German Sixth Army, cutting it off from reinforcements and supply lines. |
| The German leadership refused to allow a tactical retreat | Hitler insisted on holding Stalingrad at all costs, leading to massive German losses. |
| Air supply efforts by Germany failed to sustain the encircled troops | The Luftwaffe could not deliver enough supplies, leaving German troops starving and weakened. |
| The Soviet Red Army received reinforcements to maintain pressure on German forces | The Soviets constantly replenished their numbers, overwhelming the besieged German forces. |
| The German Sixth Army surrendered, leading to a massive loss of life | The surrender of General Paulus and his forces marked one of the worst German defeats of the war. |

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 preemptively attacked German forces before their invasion | The battle began with a German offensive; the Soviets did not strike first. |
| The battle included significant naval operations | The Volga River was used for supply transport, but there was no large-scale naval combat. |
| Germany launched a second invasion force to relieve Stalingrad | While Operation Winter Storm attempted to relieve the Sixth Army, it failed, and no second major force was sent. |
| The defending Soviet forces suffered critical ammunition shortages | Soviet supply lines remained largely functional, allowing for steady resupply. |
| The battle ended in a ceasefire agreement | The battle ended with an unconditional surrender, not a ceasefire. |
| Germany used chemical weapons during the battle | There is no historical evidence that chemical weapons were used in Stalingrad. |
| The battle lasted less than three months | The battle lasted over five months, from August 1942 to February 1943. |
| The German forces successfully withdrew a portion of their army | Hitler refused any withdrawal, leading to complete encirclement and surrender. |
| The Soviet Union used air superiority to dominate the battlefield | While Soviet air forces were present, air superiority was contested, and their main advantage was on the ground. |

Total False Positives (FP): **9**

**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 intense Luftwaffe bombing, turning Stalingrad into rubble | German air raids destroyed most of the city before the ground battle started. |
| Soviet women played a significant role in combat and logistics | Women served as snipers, medics, tank crew members, and factory workers. |
| German forces faced starvation and disease due to the siege | Lack of food and medical supplies caused thousands of deaths among trapped German soldiers. |
| Soviet forces used “hugging” tactics to neutralize German firepower | Soviets stayed close to German positions, making Luftwaffe bombings ineffective. |
| The Germans attempted a failed breakout attempt (Operation Winter Storm) | Field Marshal von Manstein launched an attempted relief operation, but it failed. |
| The Soviets used deception tactics (Maskirovka) to mislead German intelligence | Fake troop movements and false radio messages misled German forces. |
| Tens of thousands of Axis prisoners died in Soviet captivity after the battle | Of the 91,000 German prisoners taken, fewer than 6,000 survived captivity. |
| Stalingrad became a major Soviet propaganda victory | The battle was used as a symbol of Soviet resilience and heroism. |
| Soviet forces utilized river crossings to transport reinforcements and supplies | Despite German attacks, Soviet reinforcements successfully crossed the Volga River. |
| The Soviet Union encircled German forces using a pincer movement | Operation Uranus executed a successful double envelopment maneuver, trapping German forces. |

Total False Negatives (FN): **10**

***Final Results***

| **Metric** | **Value** |
| --- | --- |
| Total True Positives (TP) | **10** |
| Total False Positives (FP) | **9** |
| Total False Negatives (FN) | **10** |
| Precision | **52.63%** |
| Recall | **50.00%** |
| F1-Score | **51.24%** |

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

1. Major Issue: The LLM Predicted the Wrong Winner

* The LLM incorrectly assumed Germany won, when in reality, the Soviet Union decisively won.
* Germany lost the entire Sixth Army, leading to a catastrophic defeat.

2. Strengths in LLM’s Predictions

Correctly identified major aspects of the battle:

* Urban warfare and close-quarters combat dominated the battle.
* The Soviet counteroffensive led to a decisive encirclement.
* Germany’s extended supply lines were a key weakness.
* The Luftwaffe failed to sustain German forces.
* Winter conditions played a major role in Germany’s defeat.

3. Weaknesses in LLM’s Predictions

Missed key tactical moments:

* The Luftwaffe's initial bombardment destroyed the city but helped Soviet defenders.
* The German breakout attempt (Operation Winter Storm) was not mentioned.
* Soviet deception tactics played a significant role in the battle’s outcome.

Invented historical inaccuracies:

* No preemptive Soviet attack.
* No naval battles.
* No successful German withdrawal.

***Final Verdict***

* The LLM’s prediction was weak (~51.24% 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.