



# Dhirubhai Ambani University Technology

Formerly DA-IICT

**IT457 Cloud Computing**

## **Assignment-Amazon SageMaker**

### **Group 5**

<b>Name</b>	<b>Student ID</b>
Sneh Joshi	202201048
Sahil Chaudhari	202201171
Kshitij Patel	202201232
Neel Patel	202201494

# Model Analysis

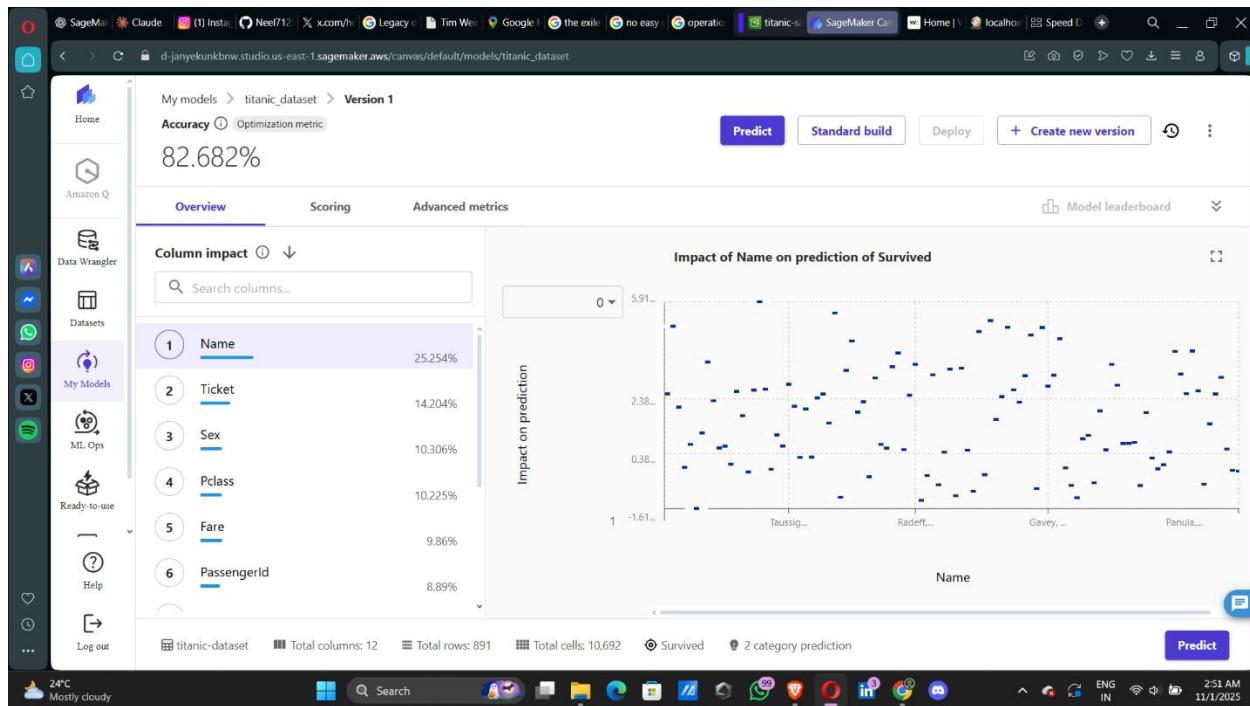
Dataset Used : [Titanic Dataset](#)

## Visualization Analysis

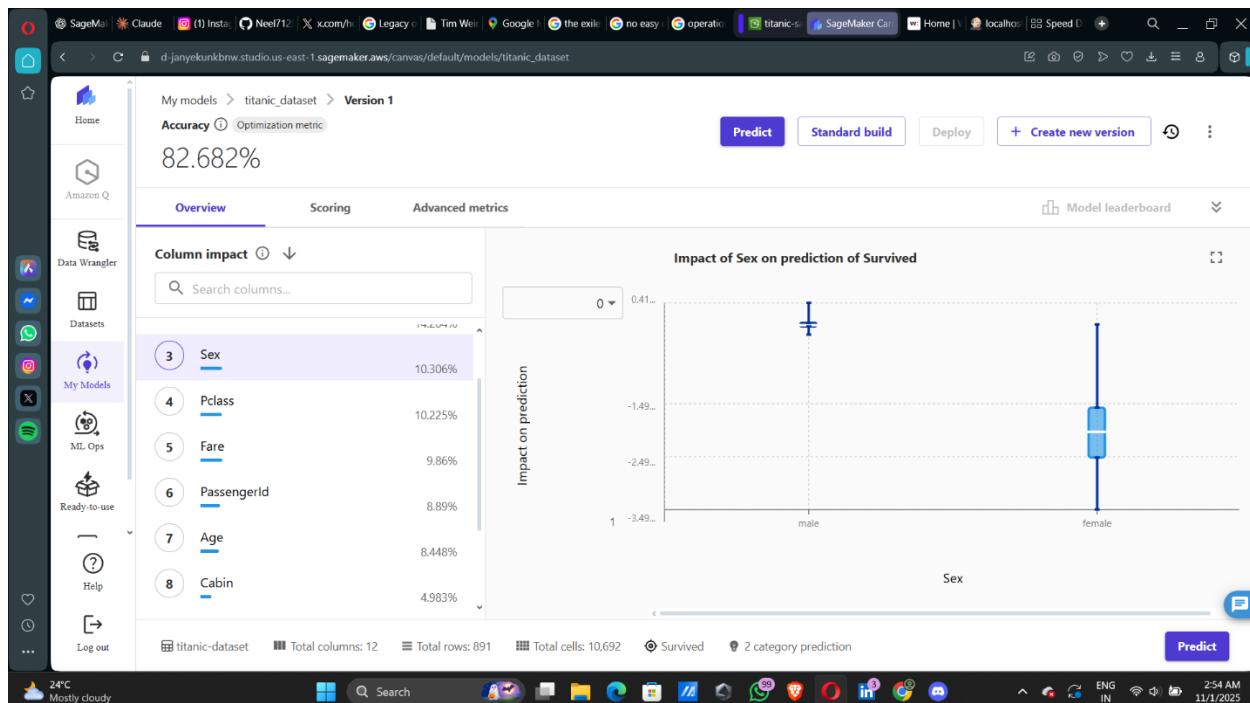
The screenshot shows the AWS SageMaker Canvas interface for the 'titanic\_dataset' model version 1. The 'Build' tab is selected. On the left, a sidebar lists various tools: Home, Amazon Q, Data Wrangler, Datasets, My Models, ML Ops, Ready-to-use, Help, and Log out. The main area has tabs for Select, Build, Analyze, Predict, and Deploy. Under 'Select a column to predict', the 'Target column' is set to 'Survived'. A 'Value distribution' chart shows a highly skewed distribution with most values at 0.00 and a few at 0.95. Under 'Model type', it's noted as '2 category prediction' for 'Survived'. A 'Configure model' button is present. Below this, a 'Data visualizer' section displays the 'titanic-dataset' with 891 rows. The table headers are Column name, Data type, Feature type, Missing, Mismatched, Unique, and Mode. The data shows columns: Ticket (Text, -), Survived (Target, Numeric, Binary, 0.00%, 0.00%, 2, 0), and SibSp (Numeric, -). A note at the bottom says 'Total columns: 12, Total rows: 891, Total cells: 10,692, Show dropped columns'. The bottom of the screen shows a Windows taskbar with various icons.

The Titanic dataset contains 891 rows and 12 columns, with “Survived” as the binary target variable (62% did not survive, 38% survived). It has **0% missing or mismatched values**, ensuring excellent data quality with no preprocessing required. Key features include **Ticket** (681 unique values, indicating shared bookings) and **SibSp** (7 unique values, mostly 0, meaning most traveled alone). The dataset’s mix of **numeric, categorical, and text** features provides a strong foundation for building an accurate survival prediction model.

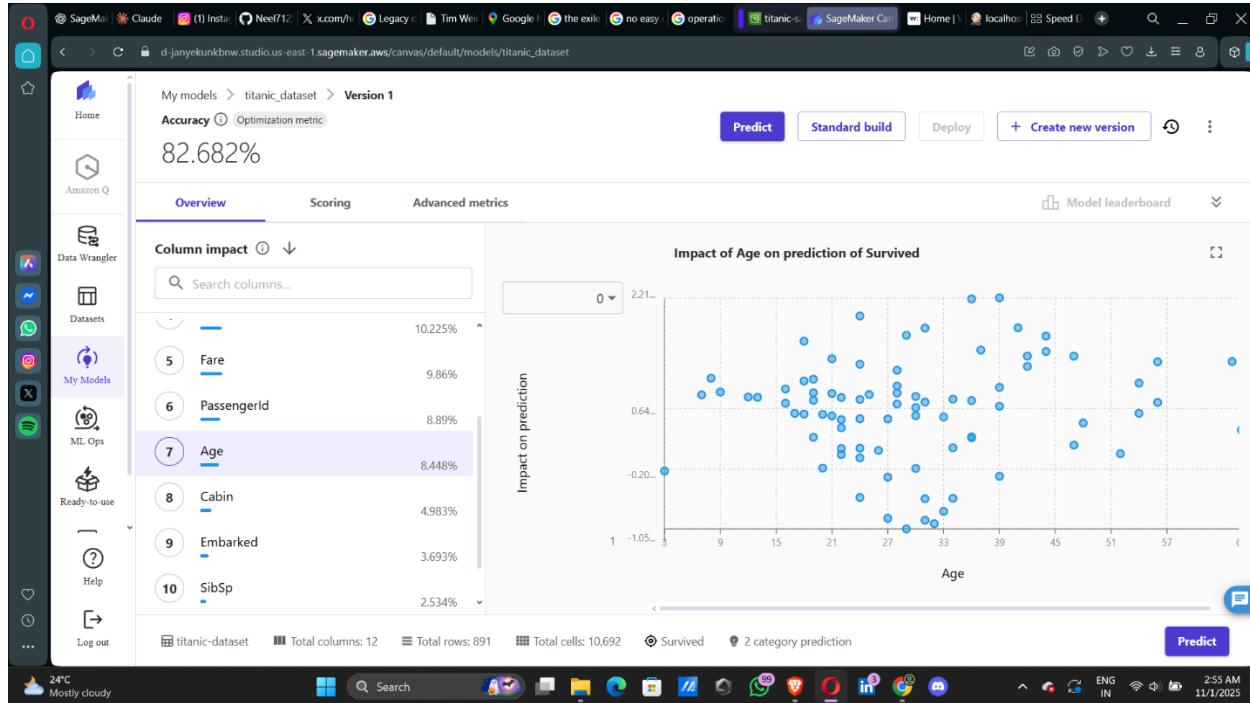
## Feature Impact Analysis



**Name (25.25%)** – The most influential feature, as titles within names (Mr., Mrs., Miss., etc.) reflected social status, gender, and age group. Titles like *Mrs.* and *Miss* strongly correlated with survival, while *Mr.* indicated lower chances.



**Sex (10.31%)** – A major predictor showing women had much higher survival impact scores, aligning with the “women and children first” rule. Female passengers were prioritized for lifeboats, making gender a key survival determinant.



**Age (8.45%)** – Younger passengers (0–15) had higher survival impacts, supporting child priority during evacuation. For ages 20–40, survival varied widely based on other factors like gender and class, showing that age interacted with social and demographic variables.

### Individual Passenger Prediction Analysis

The model analyzed a sample passenger, **Mr. Anthony Abbing**, a 24-year-old male in **third class (Pclass=3)** with no family aboard. It predicted a **98.2% chance of non-survival** and only **1.8% survival probability**. This outcome reflects several high-risk factors: being male, traveling in third class, being a young adult, and traveling alone—all historically associated with lower survival rates. The prediction tool effectively shows how combinations of **age, gender, class, and family status** influenced outcomes, offering insights into the social and demographic inequalities that shaped survival during the Titanic disaster.

SageMaker Studio - d-janyekunkbnw.studio.us-east-1.sagemaker.aws/canvas/default/models/titanic\_dataset

My models > titanic\_dataset > Version 1

Select Build Analyze Predict Deploy

Modify values to predict Survived in real time.

Filter columns

Column	Value
PassengerId	1
Pclass	3
Name	Abbing, Mr. Anthony
Sex	male
Age	24
SibSp	0

Survived Prediction

No

New prediction Last prediction

No 98.203% ⓘ

Yes 1.797% ⓘ

Download prediction

24°C Mostly cloudy

Search

2:53 AM 11/1/2025

Home Amazon Q Data Wrangler Datasets My Models ML Ops Ready-to-use Help Log out

## Accuracy

SageMaker Studio - d-janyekunkbnw.studio.us-east-1.sagemaker.aws/canvas/default/models/titanic\_dataset

My models > titanic\_dataset > Version 1

Accuracy ⓘ Optimization metric 82.682%

Predict Standard build Deploy + Create new version

Overview Scoring Advanced metrics Model leaderboard

Predicted vs. Actual

All predictions

Model accuracy insights

If the model predicts 0, it is correct 85.586% of the time. ⓘ

For the values that are 0 in the dataset, the model predicted 86.364% of them to be 0. ⓘ

titanic-dataset Total columns: 12 Total rows: 891 Total cells: 10,692 Survived 2 category prediction

Predict

24°C Mostly cloudy

Search

2:56 AM 11/1/2025

Home Amazon Q Data Wrangler Datasets My Models ML Ops Ready-to-use Help Log out

The Sankey diagram in the Scoring tab visualizes the model's strong performance across **179 test predictions** with minimal misclassifications. For passengers predicted **not to survive (Class 0)**, most correctly map to actual non-survivors, while only a few represent false negatives. Similarly, for **predicted survivors (Class 1)**, most align with actual survivors, with few false positives. The model achieves **~85.6% accuracy for non-survival predictions** and correctly identifies **~86.4% of actual non-survivors**, reflecting balanced performance. The clear, minimal crossover in the diagram confirms that the model effectively distinguishes survivors from non-survivors without major bias toward either class.

## Advanced Metrics Analysis

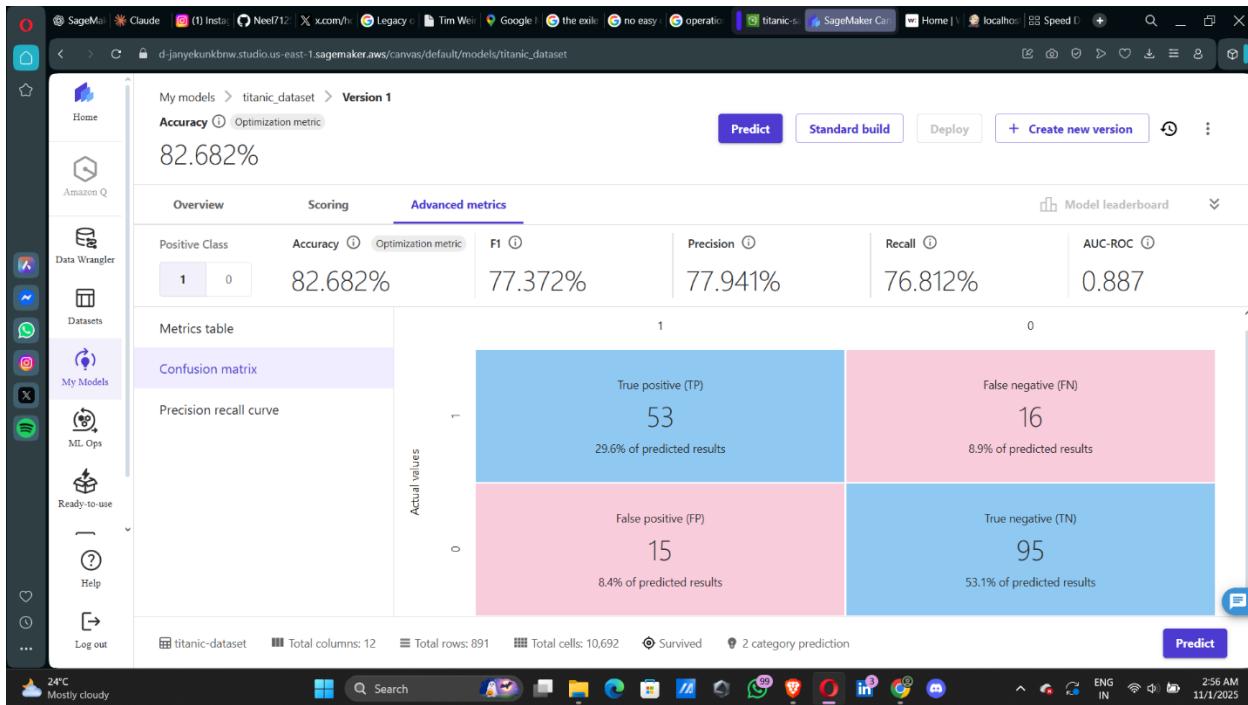
### Metrics table

Positive Class	Accuracy	F1	Precision	Recall	AUC-ROC
1	82.682%	77.372%	77.941%	76.812%	0.887

Metric name	Value
precision	0.779
recall	0.768
accuracy	0.827
f1	0.774
auc	0.887

With **Precision = 0.779** and **Recall = 0.768**, the model accurately identifies most survivors while maintaining few false alarms. The **F1 Score (0.774)** reflects strong balance, and the **AUC (0.887)** confirms excellent separation between survival outcomes.

## Confusion matrix



The confusion matrix breaks down the model's 179 test predictions into four categories: 53 True Positives (29.6% of predictions) where the model correctly identified survivors, 95 True Negatives (53.1%) where non-survivors were correctly predicted, 16 False Negatives (8.9%) representing survivors incorrectly classified as non-survivors, and 15 False Positives (8.4%) representing non-survivors incorrectly predicted to survive. The near-equal error rates (8.9% vs 8.4%) demonstrate the model's balanced performance without significant bias toward either class, while the dominant blue quadrants (TP and TN) visually confirm that correct predictions far outnumber errors.

## Precision recall curve

The **Precision-Recall curve (AUPRC = 0.87)** shows that the model maintains **high precision (>0.8)** for recall values up to about 0.5, then gradually declines as recall increases. This indicates strong confidence in positive predictions with few false positives. The **AUPRC far above the baseline ( $\approx 0.38$ )** confirms the model's effectiveness in identifying real survivors. The smooth curve reflects **stable and reliable performance**, suggesting good generalization to unseen Titanic data.

SageMaker Claude Insta Neel712 x.com Legacy Tim Weil Google the exile no easy operation titanic-s SageMaker Canvas Home localhost Speed

My models > titanic\_dataset > Version 1

Accuracy (Optimization metric) 82.682%

Predict Standard build Deploy + Create new version

Overview Scoring Advanced metrics Model leaderboard

Positive Class	Accuracy (Optimization metric)	F1 (Optimization metric)	Precision (Optimization metric)	Recall (Optimization metric)	AUC-ROC (Optimization metric)
1 0	82.682%	77.372%	77.941%	76.812%	0.887

Metrics table

Confusion matrix

Precision recall curve

Precision vs. recall

Precision-recall (AUPRC = 0.87)

Precision

Recall

titanic-dataset Total columns: 12 Total rows: 891 Survived 2 category prediction

Predict

24°C Mostly cloudy

2:56 AM 11/1/2025

Home Amazon Q Data Wrangler Datasets My Models ML Ops Ready-to-use Help Log out