| **Graph** | **What it shows** | **File name** |
| --- | --- | --- |
| **Accuracy vs AUC Bar Plot** | Compare model performance across companies | xgboost\_accuracy\_auc\_comparison.png |
| **ROC Curves for All Companies** | Visualize classification quality (per company) | xgboost\_roc\_curves\_all\_companies.png |
| **Feature Importance Plots** | Top features driving prediction for each company | (Inside Excel: xgboost\_feature\_importances.xlsx) |

**1. Accuracy vs AUC Comparison Bar Plot**

✅ **Purpose**:  
Compare how well XGBoost models work for each company.

✅ **How to interpret**:

* **Accuracy** = % of correct predictions (Up/Down correctly predicted)
* **AUC** = How well model separates Up vs Down cases (AUC closer to 1 = better)

✅ **Read the plot**:

* Look for companies with **high Accuracy** **AND** **high AUC**.
* If a company has Accuracy > 60% and AUC > 0.65 → Good predictive performance!
* If Accuracy is low (<55%) and AUC close to 0.5 → model is almost random guessing.

✅ **Example**:  
If AAPL has Accuracy 68% and AUC 0.72 →  
→ News + TimeBlock can **meaningfully predict** AAPL stock moves.

**2. ROC Curves for All Companies**

✅ **Purpose**:  
Visually see how "good" the model is for **each company**.

✅ **How to interpret**:

* **X-axis** = False Positive Rate (wrong alarms)
* **Y-axis** = True Positive Rate (correct detection)
* **Diagonal line** = random guessing (no skill)

✅ **Read the plot**:

* **Curves above the diagonal** → model is useful!
* **Higher curve** = better model.
* **Closer to top-left corner** = best classification (perfect).

✅ **Example**:

* AMD’s ROC curve much higher than TSLA’s →  
  Model can better distinguish Up/Down for AMD than for TSLA.

✅ **Area Under Curve (AUC)** helps summarize:

* AUC > 0.7 → Good model
* AUC ~0.5 → Random guessing

**3. Feature Importance Plots (and Excel Table)**

✅ **Purpose**:  
Find out which **FinBERT embeddings** (and Time Block) are **most important** for predicting Up/Down.

✅ **How to interpret**:

* **Higher importance score** → feature plays bigger role in model.
* See which FinBERT variable (latent topic) has the strongest impact.
* If **Time\_Block\_Count** is important → stock movement depends not only on news but also on time progression.

✅ **Read the table**:

* Each company has different top 5 features.
* If **finbert\_59** is top across many companies → that news topic is consistently important!

✅ **Example**:

* For AAPL, finbert\_59 and finbert\_48 are top features →  
  News patterns related to those latent dimensions **strongly drive AAPL stock moves**.