

Staci McDuffie
Bellevue University
DSC 680 – Applied Data Science
Professor Iranitalab
November 07, 2025

Figure 1: AUC Bar Chart

```
models = ["Logistic Regression", "Random Forest", "XGBoost"]
aucs = [0.82, 0.89, 0.91]
fig1, ax1 = plt.subplots()
ax1.bar(models, aucs)
ax1.set_ylim(0.7, 1.0)
ax1.set_ylabel("Mean ROC AUC")
ax1.set_title("Figure 1. Cross-validated ROC AUC by Model")
for i, v in enumerate(aucs):
    ax1.text(i, v + 0.005, f'{v:.2f}', ha="center", va="bottom", fontsize=8)
pdf.savefig(fig1)
plt.close(fig1)
```

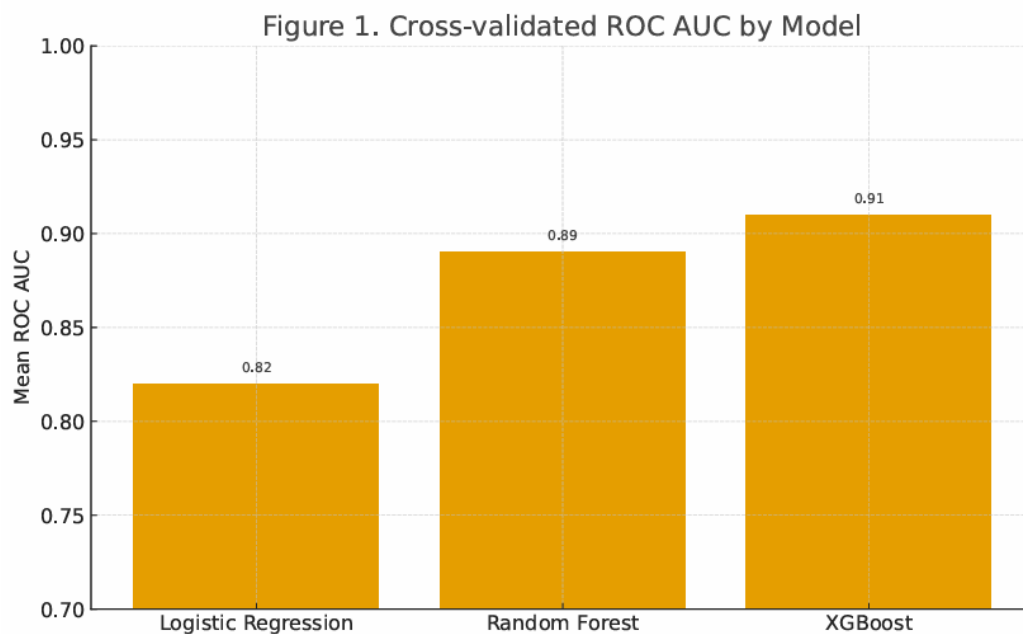


Figure 2: ROC Curve

```
fpr = np.array([0.0, 0.02, 0.05, 0.10, 0.18, 0.30, 1.0])
tpr = np.array([0.0, 0.40, 0.65, 0.78, 0.88, 0.93, 1.0])
fig2, ax2 = plt.subplots()
ax2.plot(fpr, tpr, marker="o", label="XGBoost ROC")
ax2.plot([0, 1], [0, 1], linestyle="--", label="No skill")
ax2.set_xlabel("False Positive Rate (1 - Specificity)")
ax2.set_ylabel("True Positive Rate (Sensitivity)")
ax2.set_title("Figure 2. ROC Curve – Tuned XGBoost Model (Simulated)")
ax2.legend()
pdf.savefig(fig2)
plt.close(fig2)
```

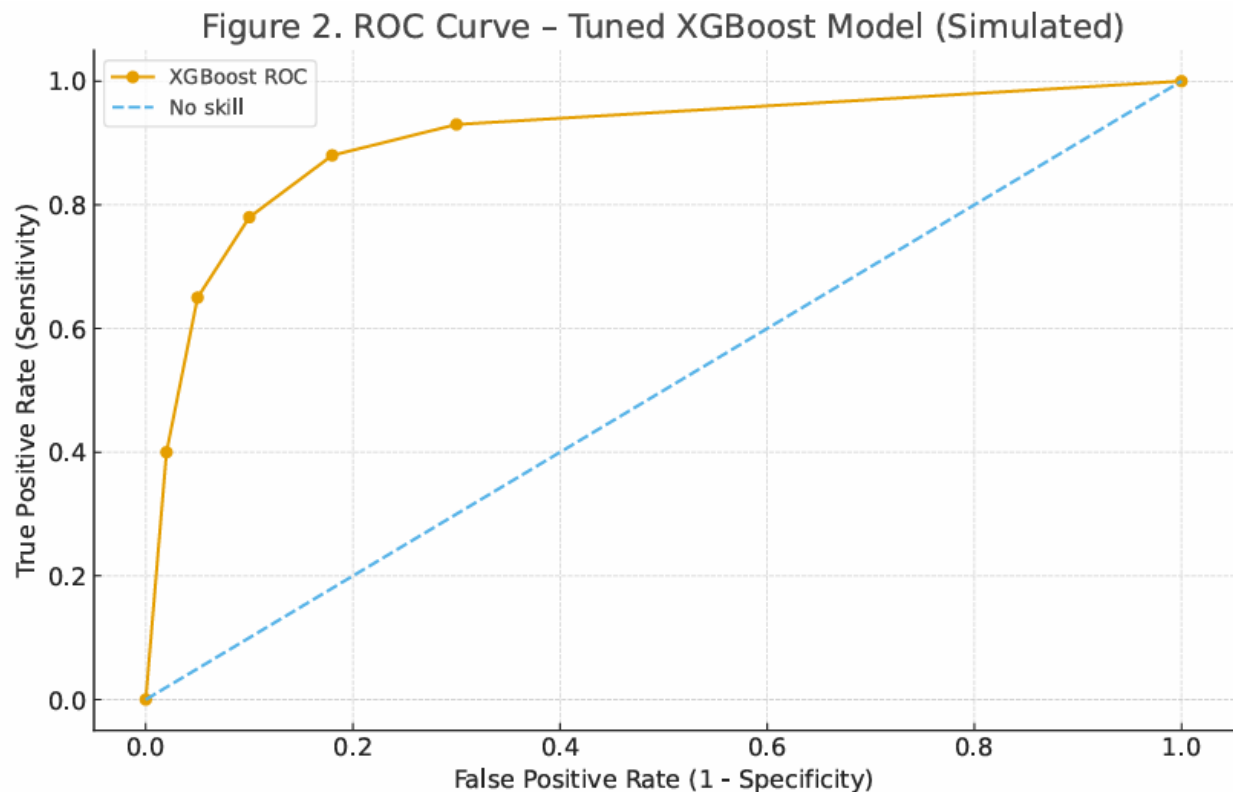


Figure 3: Variable Importance

```
features = ["Smoking", "Alcohol", "Age", "Accident/Trauma", "Fever Last Year",  
           "Hours Sitting", "Season", "Childhood Diseases"]  
importance = [0.25, 0.20, 0.18, 0.10, 0.08, 0.07, 0.06, 0.06]  
fig3, ax3 = plt.subplots()  
y_pos = np.arange(len(features))  
ax3.barh(y_pos, importance)  
ax3.set_yticks(y_pos)  
ax3.set_yticklabels(features)  
ax3.invert_yaxis()  
ax3.set_xlabel("Relative Importance")  
ax3.set_title("Figure 3. Variable Importance – XGBoost Model (Simulated)")  
pdf.savefig(fig3)  
plt.close(fig3)
```

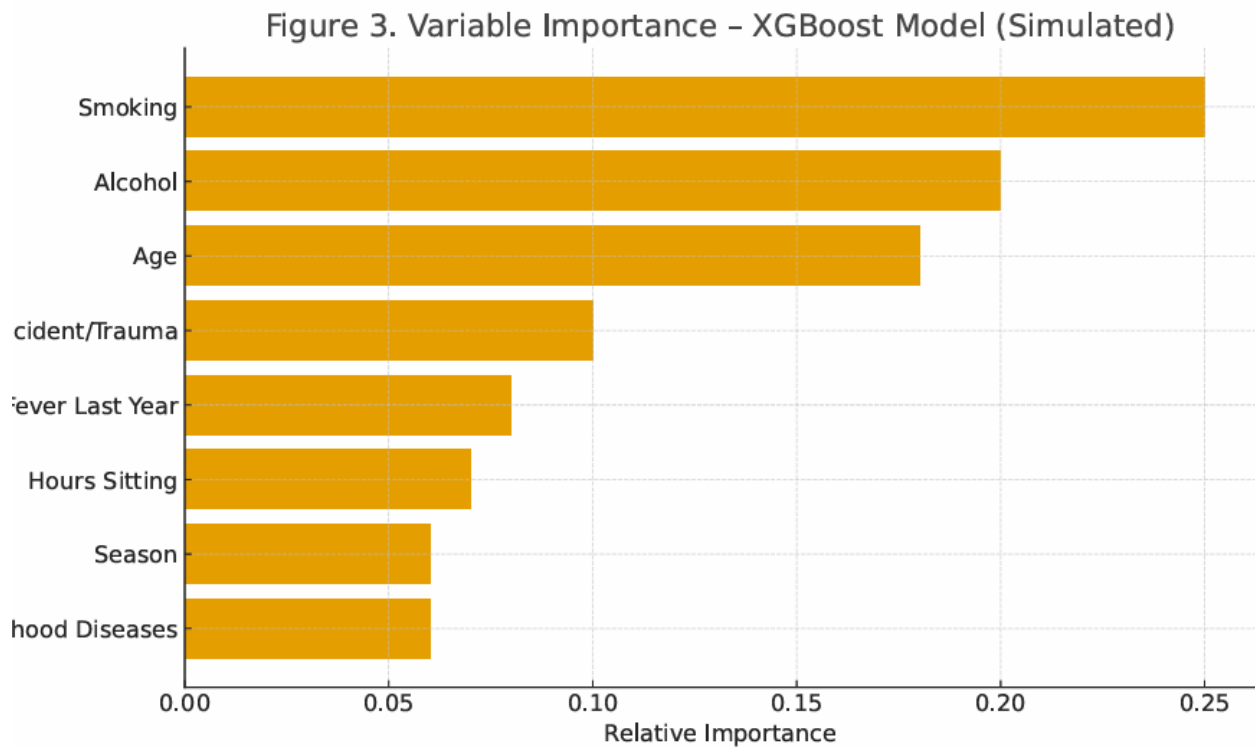


Figure 4: Confusion Matrix Heatmap

```
cm = np.array([[22, 2],
               [3, 7]]) # rows: true normal/alterd; cols: predicted normal/alterd

fig4, ax4 = plt.subplots()
cax = ax4.imshow(cm)
ax4.set_xticks([0, 1])
ax4.set_yticks([0, 1])
ax4.set_xticklabels(["Pred Normal", "Pred Altered"])
ax4.set_yticklabels(["True Normal", "True Altered"])
ax4.set_title("Figure 4. Confusion Matrix – XGBoost Model (Simulated)")

for i in range(cm.shape[0]):
    for j in range(cm.shape[1]):
        ax4.text(j, i, str(cm[i, j]), ha="center", va="center", color="w", fontsize=10)

fig4.colorbar = plt.colorbar(cax, ax=ax4)
pdf.savefig(fig4)
plt.close(fig4)
```

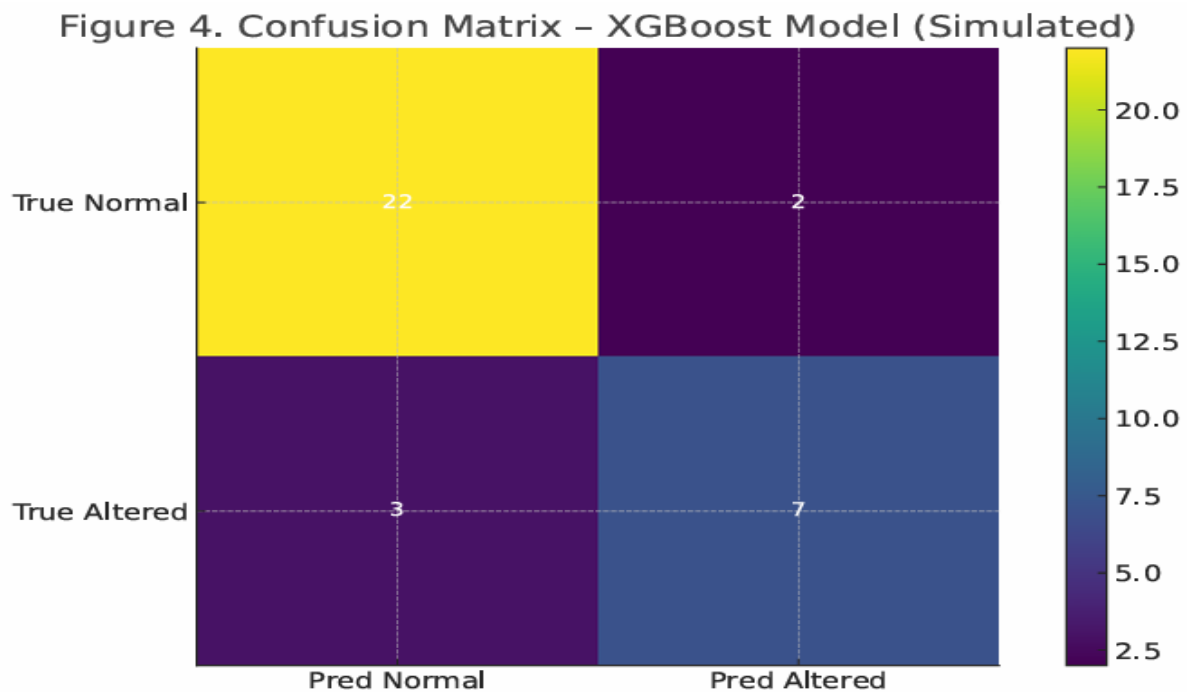


Figure 5: Fertility Status by Smoking Habit

```
smoking_levels = ["Low", "Moderate", "High"]
normal_counts = np.array([40, 30, 18])
altered_counts = np.array([2, 4, 6])
total = normal_counts + altered_counts
normal_prop = normal_counts / total
altered_prop = altered_counts / total
x = np.arange(len(smoking_levels))

fig5, ax5 = plt.subplots()
ax5.bar(x, normal_prop, label="Normal")
ax5.bar(x, altered_prop, bottom=normal_prop, label="Altered")
ax5.set_xticks(x)
ax5.set_xticklabels(smoking_levels)
ax5.set_ylabel("Proportion of Cases")
ax5.set_ylim(0, 1)
ax5.set_title("Figure 5. Fertility Status by Smoking Habit (Simulated)")
ax5.legend()
pdf.savefig(fig5)
plt.close(fig5)
```

Figure 5. Fertility Status by Smoking Habit (Simulated)

