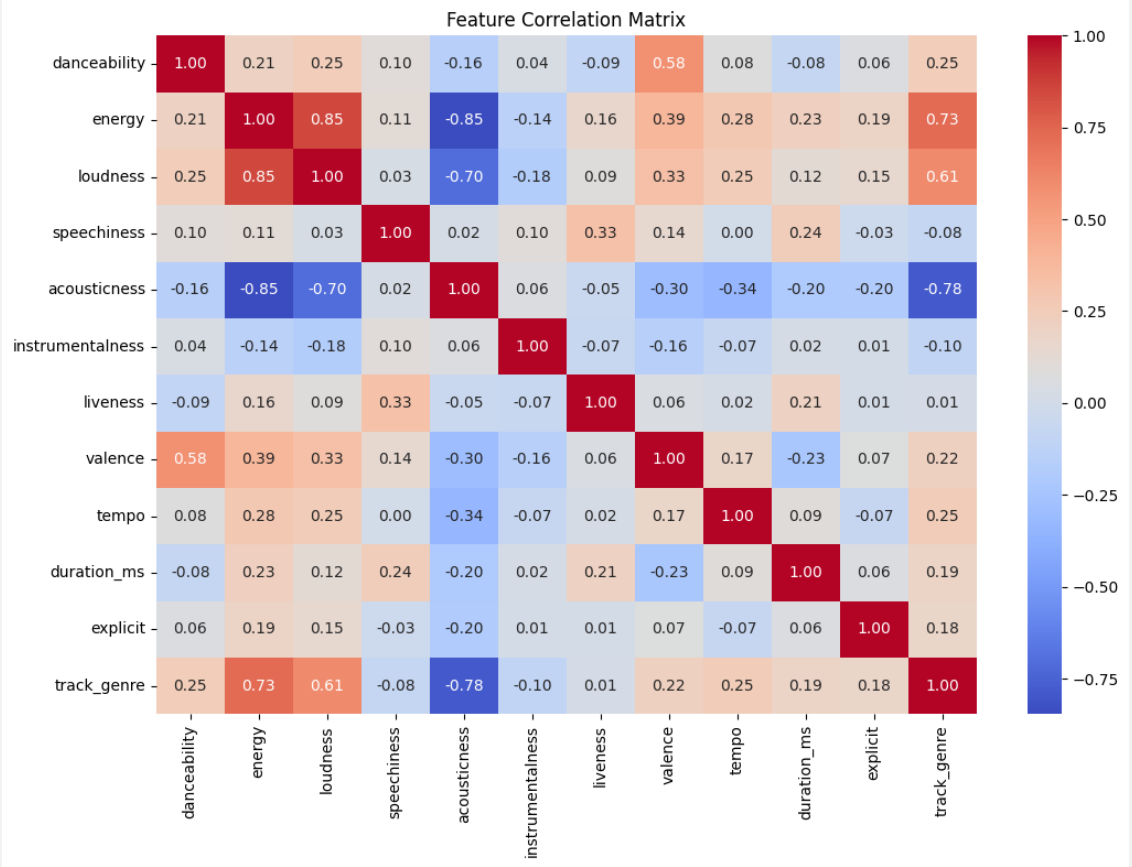
# REMOVING CORRELATION ANOMALIES TO ACHIEVE A HIGHEST LEVEL OF ACCURACY

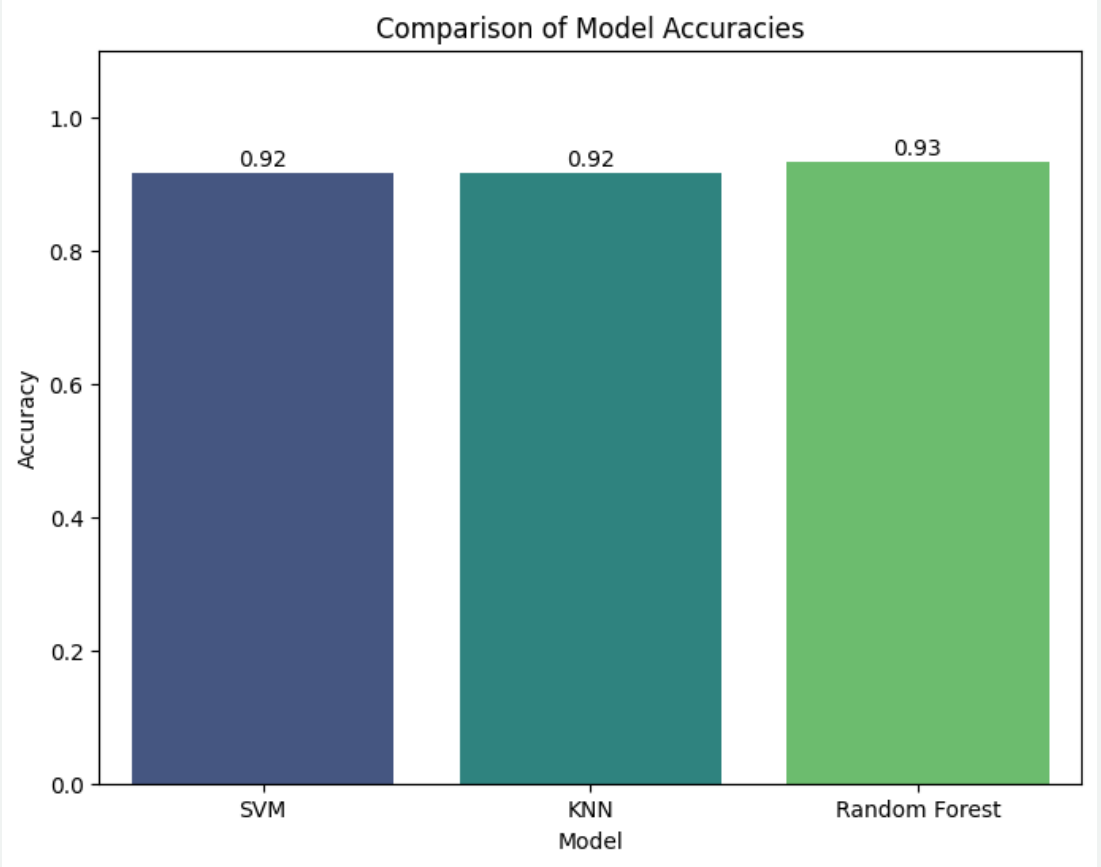
Jose Carlo Burga

TOP CORRELATION ANOMALIES:
- ACOUSTICNESS
- ENERGY
- LOUDNESS

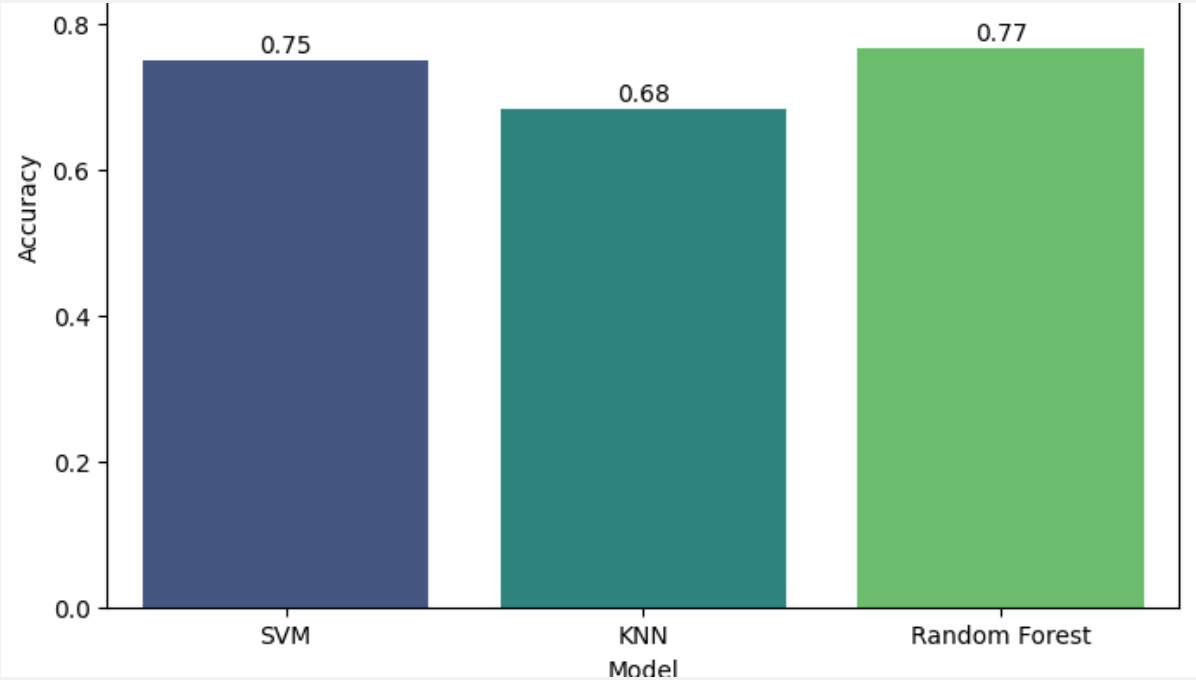
# Feature Correlation Matrix



# EXPERIMENT 1: WHAT IF I DON'T REMOVE ANY ANOMALIES

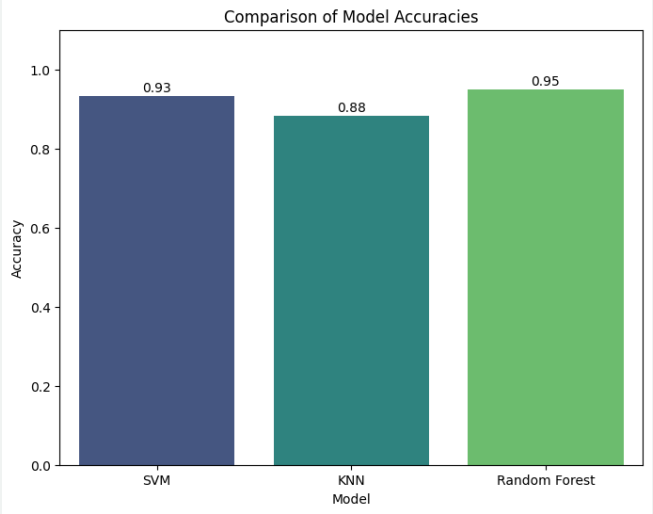


# EXPERIMENT 2: WHAT IF I REMOVE THE TOP THREE ANOMALIES

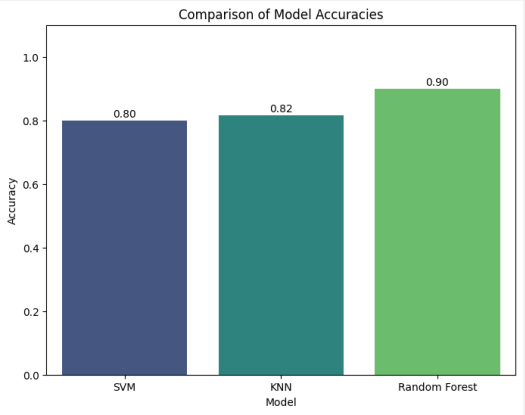


# EXPERIMENT 3: WHAT IF I REMOVE TWO ANOMALIES AT A TIME

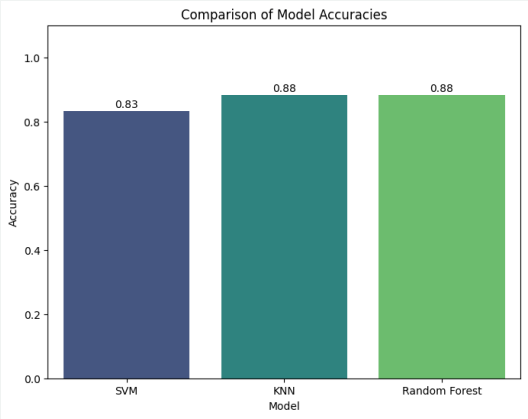
## Energy & Loudness




## Energy & Acousticness

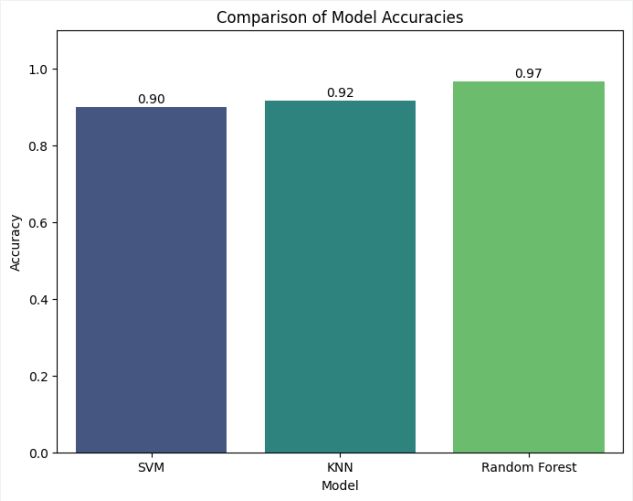



## Loudness & Acousticness

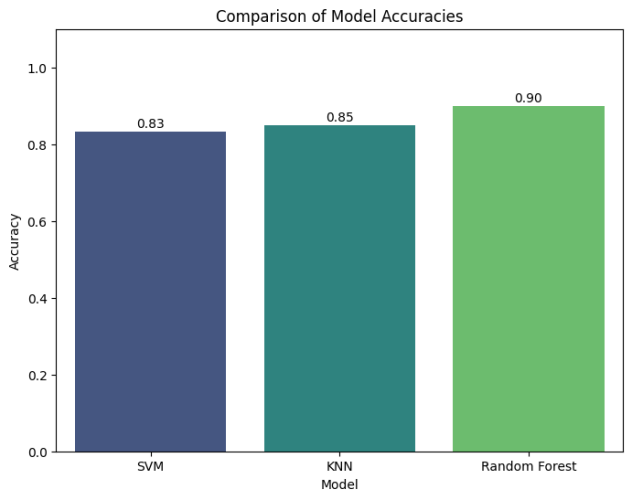


# EXPERIMENT 4: WHAT IF I REMOVE ONE ANOMALY AT A TIME

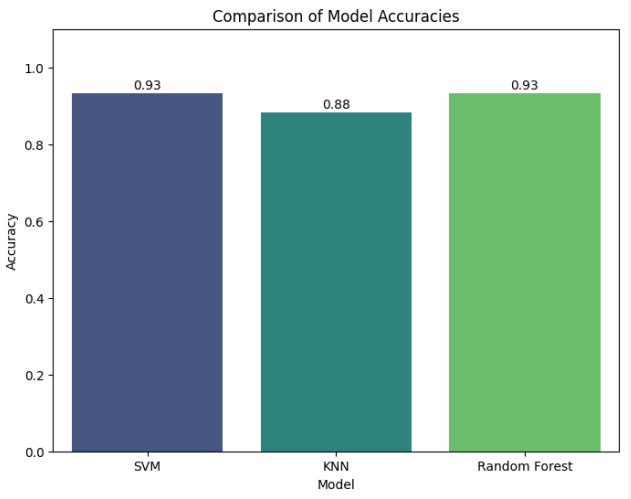
## Energy




## Acousticness

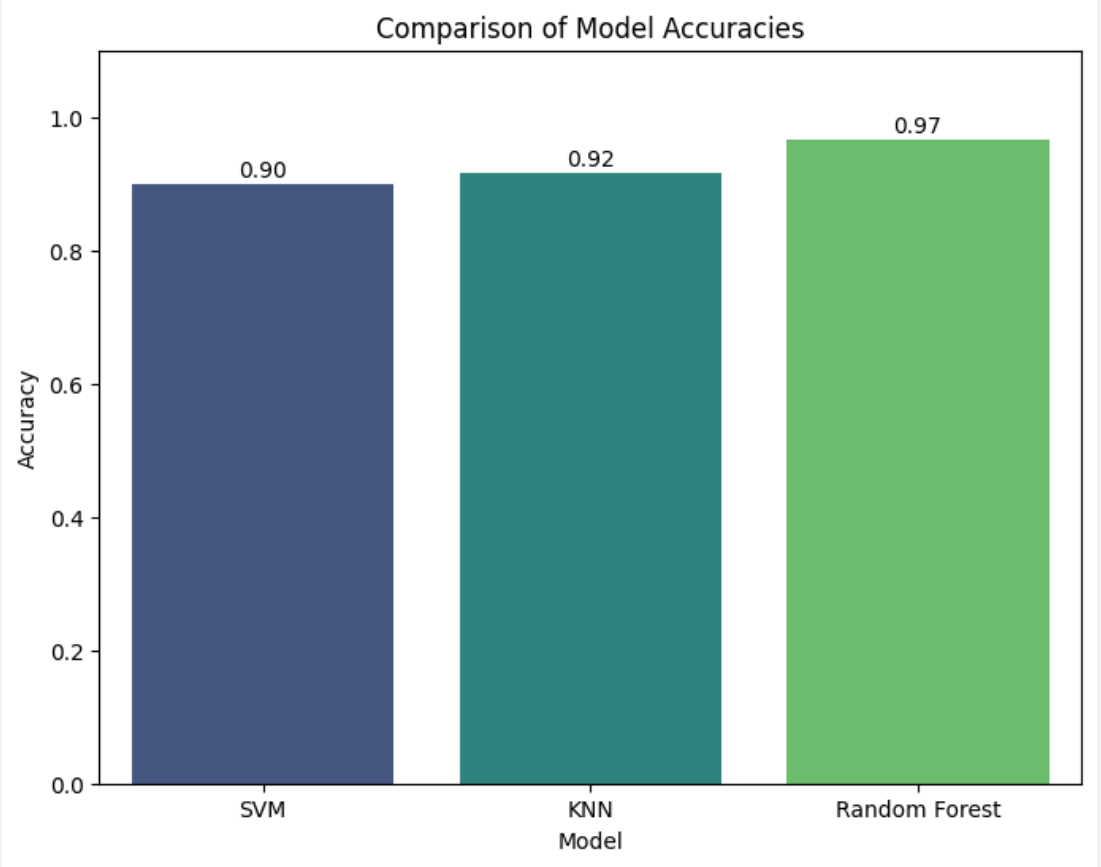



## Loudness



MODEL WITH THE HIGHEST LEVEL OF ACCURACY ACROSS EXPERIMENTS:
RANDOM FOREST:

REMOVING CORRELATION ANOMALIES




Experiment 2: Removing 'energy'
- Remove one highly correlated feature with a correlation around |0.7| or higher.
- Remove and then re-train your models.
- Note any differences compared to the baseline.

# MODEL WITH THE HIGHEST LEVEL OF ACCURACY ACROSS EXPERIMENTS: RANDOM FOREST:

# REMOVING CORRELATION ANOMALIES

Random forest run the highest level of accuracy across 4 different experiments testing the 3 highest correlated features ('energy', 'loudness', and 'acousticness')

1. Experiment 1: Keeping the three correlation anomalies
2. Experiment 2: Removing the three correlation anomalies
3. Experiment 3A: Removing two correlation anomalies at a time ('energy', 'loudness')
4. Experiment 3B: Removing two correlation anomalies at a time ('energy' and 'acousticness')
5. Experiment 3C: Removing two correlation anomalies at a time ( 'loudness' and 'acousticness')
6. Experiment 4A: Removing one correlation anomaly at a time ('acousticness')
7. Experiment 4B: Removing one correlation anomaly at a time ( 'energy')
8. Experiment 4C: Removing one correlation anomaly at a time ( 'loudness')

Removing correlation anomalies to achieve a highest level of accuracy

Click here for the experiment notebooks in GitHub

Experiment 2: Removing 'energy'
- Remove one highly correlated feature with a correlation around |0.7| or higher.
- Remove and then re-train your models.
- Note any differences compared to the baseline.