



Analyzing Taylor Swift's Songs Features and Emotion Prediction

**HARNAIN KAUR SARDARNI
BONUS PROJECT**



The dataset provides us a comprehensive overview of Taylor Swift's songs, encompassing both its quantitative and qualitative attributes. It provides detailed features of Taylor Swift's songs, including danceability, energy, and valence. It also includes the track information such as release date and it's popularity.

The project aimed to analyze Taylor Swift's music using quantitative features and machine learning techniques to gain insights into her songs “emotional characteristics”. By categorizing her songs into different emotional states and exploring factors influencing these states, I aimed to deepen our understanding of her music's impact on listeners and shed light on patterns in her discography. Additionally, the project sought to showcase the potential of data driven analysis in music research and provide us a practical application of machine learning in the realm of music analysis.

Data Preprocessing:

Key Features Statistics:

- Looked into features like Mean, Std, Min, Max for features like danceability, energy, loudness, etc.
- Highlights the average values to provide us insights into the Taylor Swift's music style.

No Missing Values:

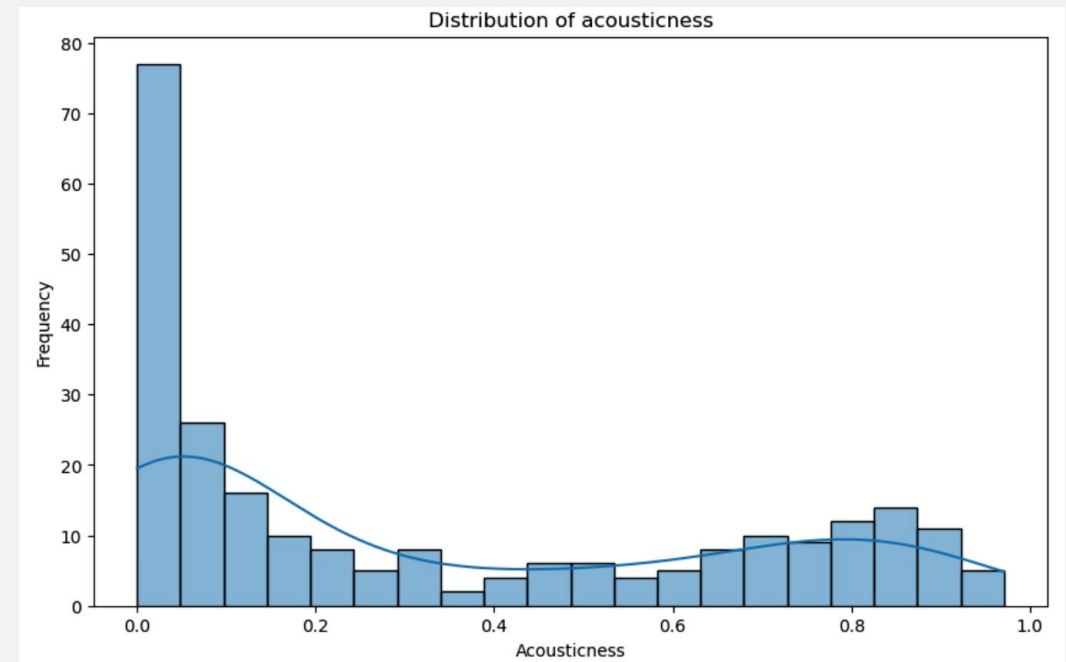
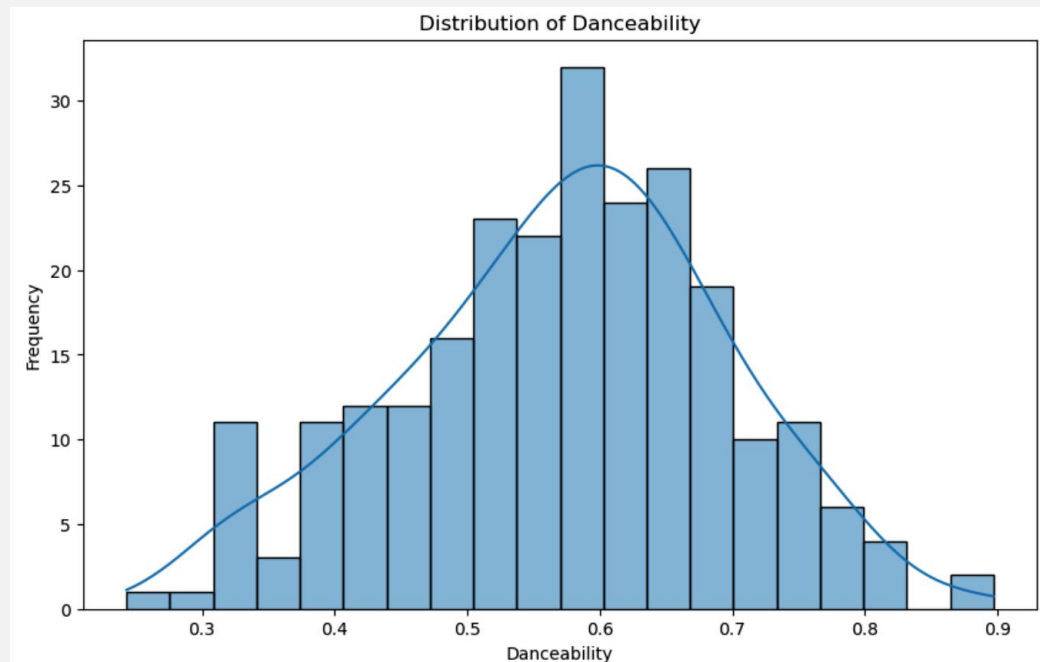
- By looking at the image we can confirm the absence of missing data in the dataset.

```
: taylor.isnull().sum()
: Unnamed: 0      0
  album_id        0
  album_name      0
  id              0
  track_name      0
  danceability    0
  energy          0
  key             0
  loudness        0
  mode            0
  speechiness     0
  acousticness    0
  instrumentality 0
  liveness        0
  valence         0
  tempo           0
  dtype: int64
```

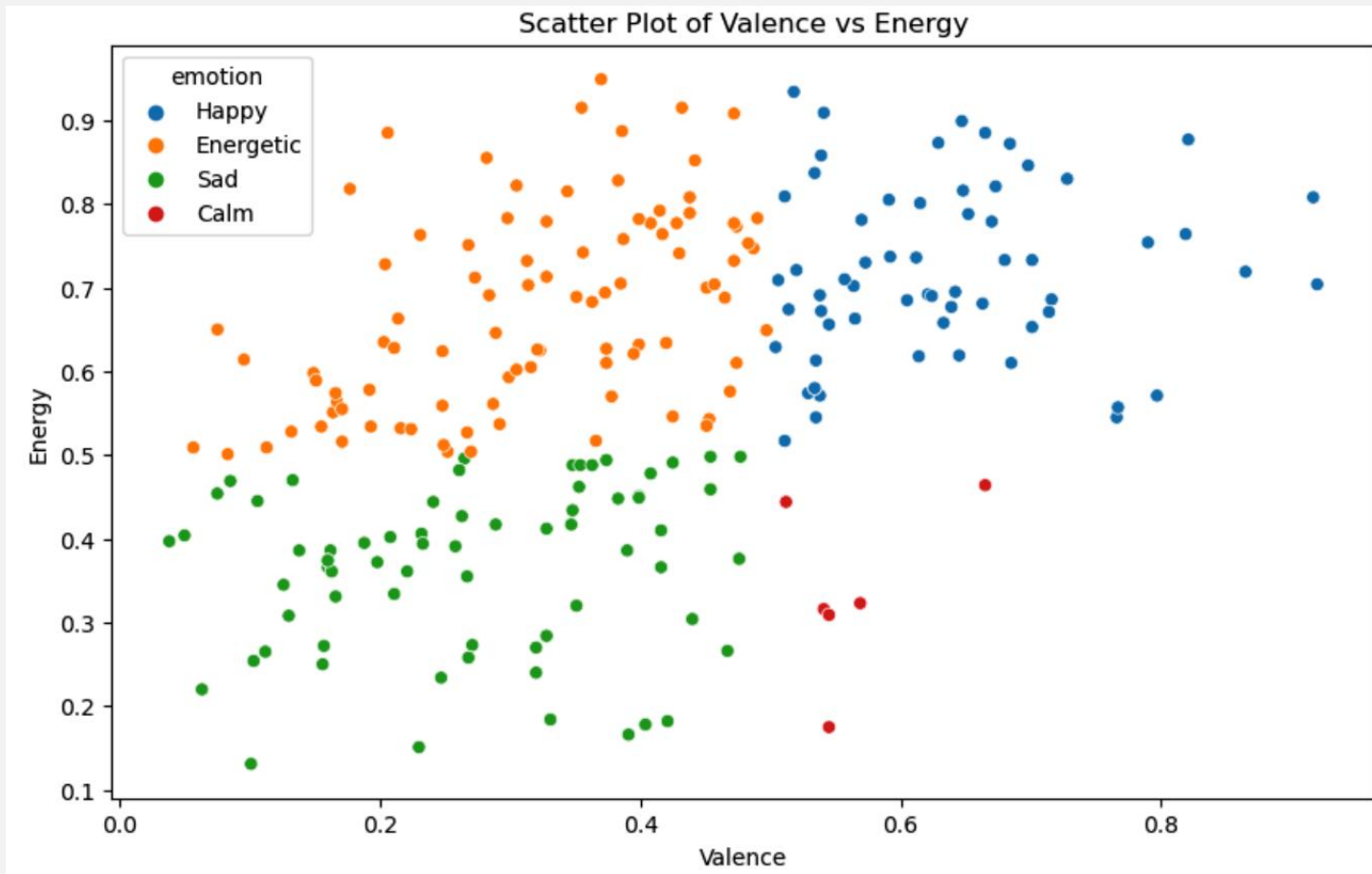
Data Visualization:

Histogram - Danceability: This visualization depicts the distribution of danceability in Taylor Swift's songs, showcasing the suitability for dancing and indicating consistency in tempo and rhythm.

Histogram - Acousticness: This histogram illustrates the distribution of acousticness, revealing the frequency of acoustic elements in Taylor Swift's music and highlighting trends in her production style.



Scatter Plot - Valence vs Energy: The scatter plot displays the relationship between valence and energy, color-coded by emotion classification, offering insights into the emotional tone and mood variation across Taylor Swift's discography.



Model Training and Evaluation:

Logistic Regression:

- Precision, recall, and F1-score for each emotion category.
- Accuracy: 92%.

Random Forest:

- Precision, recall, and F1-score for each emotion category.
- Accuracy: 98%.

Neural Network:

- Precision, recall, and F1-score for each emotion category.
- Accuracy: 76%.

Logistic Regression Report:				
	precision	recall	f1-score	support
Calm	0.00	0.00	0.00	1
Energetic	0.92	0.96	0.94	23
Happy	1.00	1.00	1.00	14
Sad	0.83	0.83	0.83	12
accuracy			0.92	50
macro avg	0.69	0.70	0.69	50
weighted avg	0.90	0.92	0.91	50

Accuracy: 0.92

Random Forest Report:				
	precision	recall	f1-score	support
Calm	1.00	1.00	1.00	1
Energetic	1.00	0.96	0.98	23
Happy	1.00	1.00	1.00	14
Sad	0.92	1.00	0.96	12
accuracy			0.98	50
macro avg	0.98	0.99	0.98	50
weighted avg	0.98	0.98	0.98	50

Accuracy: 0.98

Neural Network Report:				
	precision	recall	f1-score	support
Calm	0.00	0.00	0.00	1
Energetic	0.81	0.74	0.77	23
Happy	0.91	0.71	0.80	14
Sad	0.61	0.92	0.73	12
accuracy			0.76	50
macro avg	0.58	0.59	0.58	50
weighted avg	0.77	0.76	0.76	50

Accuracy: 0.76

Improving Model Performance:

Cross Validation for Logistic Regression:

- Cross-validation scores: [0.90, 0.94, 0.82, 0.94, 0.80], Mean CV score: 0.88

Hyperparameter Tuning for Random Forest:

- Best parameters found: max_depth: 10, min_samples_leaf: 1, min_samples_split: 2, n_estimators: 100

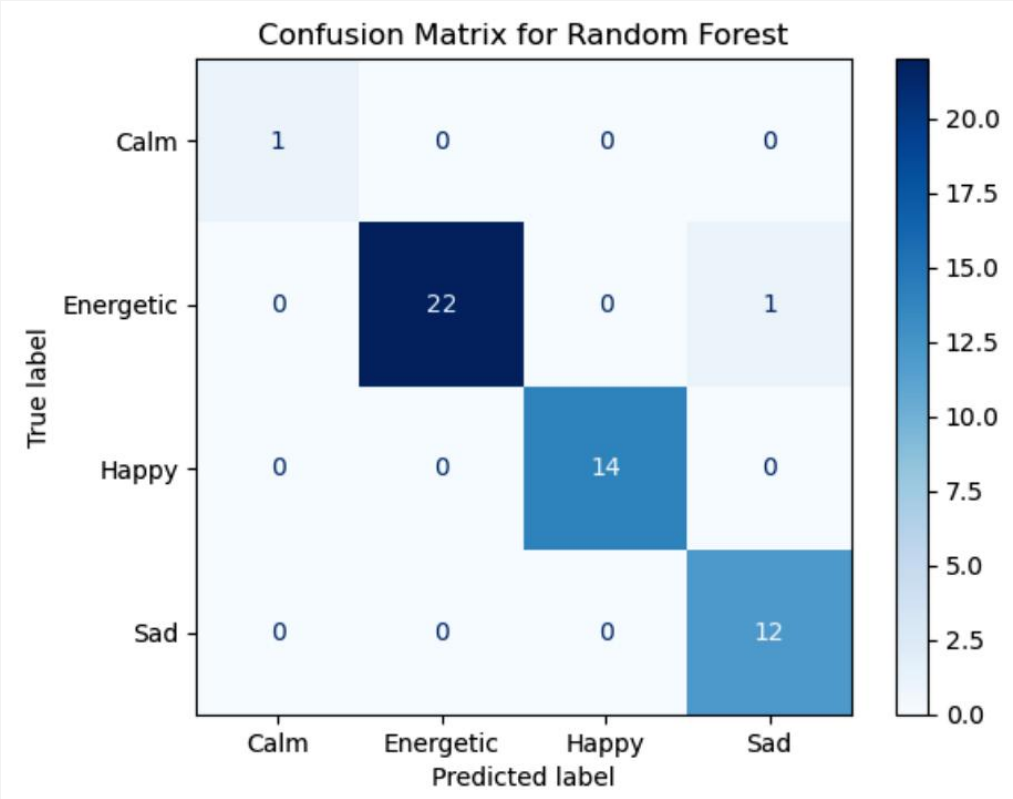
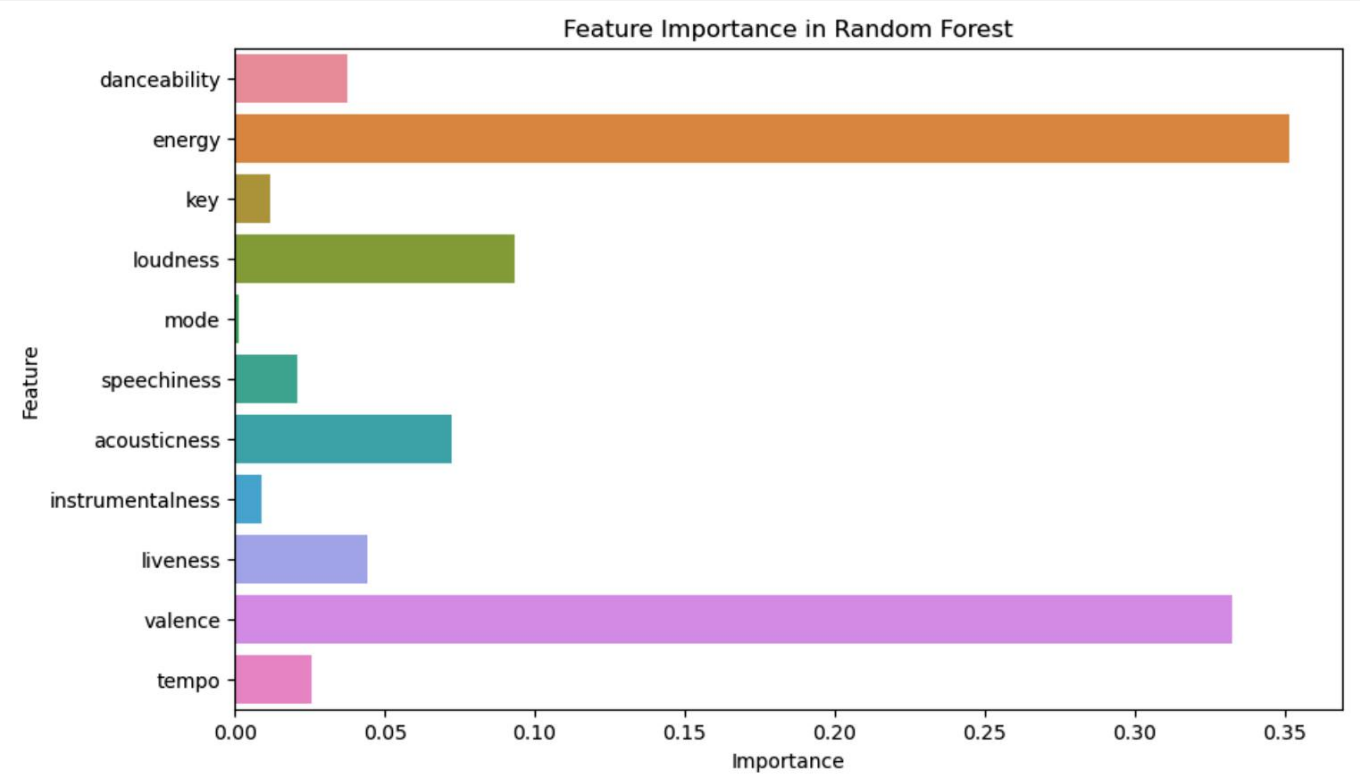
Tuned Random Forest Model:

- Precision, recall, and F1-score for each emotion category, Accuracy improved to 96%.

Tuned Random Forest Report:				
	precision	recall	f1-score	support
Calm	0.00	0.00	0.00	1
Energetic	1.00	0.96	0.98	23
Happy	1.00	1.00	1.00	14
Sad	0.86	1.00	0.92	12
accuracy			0.96	50
macro avg	0.71	0.74	0.73	50
weighted avg	0.95	0.96	0.95	50
Accuracy: 0.96				

Utilizing cross-validation and hyperparameter tuning techniques improved the model's performance, particularly in predicting 'Calm' emotions.

Further Visualizations:



Emotion Prediction Function

```
emotion_distribution = taylor['emotion'].value_counts()
print(emotion_distribution)
```

```
emotion
Energetic    96
Sad          81
Happy        63
Calm         6
Name: count, dtype: int64
```

```
def predict_emotion(song_features):
    """
    Predict the emotion of a song based on its features.

    Parameters:
    song_features (pd.DataFrame): A DataFrame containing the features of the song.

    Returns:
    str: The predicted emotion.
    """
    song_features_scaled = scaler.transform(song_features)
    predicted_emotion = best_rf.predict(song_features_scaled)
    return predicted_emotion[0]
```

```
song_indices = [106, 101, 105, 72, 13]
```

```
for song_index in song_indices:
    song_features = taylor.loc[[song_index], feature_names]
    song_name = taylor.loc[song_index, 'track_name']
    actual_emotion = taylor.loc[song_index, 'emotion']
    predicted_emotion = predict_emotion(song_features)
    print(f"Songs Name: {song_name}")
    print(f"Actual Emotion: {actual_emotion}")
    print(f"Predicted Emotion: {predicted_emotion}\n")
```

```
Songs Name: happiness
Actual Emotion: Sad
Predicted Emotion: Sad
```

```
Songs Name: champagne problems
Actual Emotion: Sad
Predicted Emotion: Sad
```

```
Songs Name: no body, no crime (feat. HAIM)
Actual Emotion: Happy
Predicted Emotion: Happy
```

```
Songs Name: The Lucky One (Taylor's Version)
Actual Emotion: Happy
Predicted Emotion: Happy
```

```
Songs Name: Wonderland (Taylor's Version)
Actual Emotion: Energetic
Predicted Emotion: Energetic
```



Thank you