How Fire is Your Mixtape?

Using Social Media Metrics to Predict the Growth of Track Sales after Album Release

Pedro Garzon, Vinson Luo, Reynis Vazquez

Dataset

Features include social media metrics: Facebook, Twitter, Wikipedia, Last.fm, Radiowaves Sirius, Myspace

Output Variable: iTunes track growth one week after release

Statistics for growth variable:

- Mean: 0.185
- Median: 0.079
- Standard Deviation: 0.360

Insights

- Large variance in metric for same amount of track sale growth
- Social media buzz actually hurts sales for most artists (negative correlation of features)
- Sales acceleration best predictor for linear regression and naive bayes

Results

<table>
<thead>
<tr>
<th>Linear Regression</th>
<th>Naïve Bayes</th>
<th>SVR 1</th>
<th>SVR 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Standard Error</td>
<td>0.3205</td>
<td>0.331</td>
<td>0.5521</td>
</tr>
<tr>
<td># Data Points</td>
<td>1543</td>
<td>1543</td>
<td>835</td>
</tr>
<tr>
<td>Features Used</td>
<td>Sales acceleration, LastFm, Twitter</td>
<td>Facebook, Twitter, Myspace</td>
<td>Facebook, Radio Sirius, LastFm*</td>
</tr>
</tbody>
</table>

* Using K-NN Interpolation

Method

- Processed data to look at only album releases
- Trained linear regression, naive bayes, and support vector regression models on data using different combinations of features
- Results shown use best combinations of features as determined by cross validation

Challenges

- Too many missing values in features
- Varying amount of features per album
- Severe overfitting of some models
- Large bias for best scenarios