

How Fire is Your Mixtape?

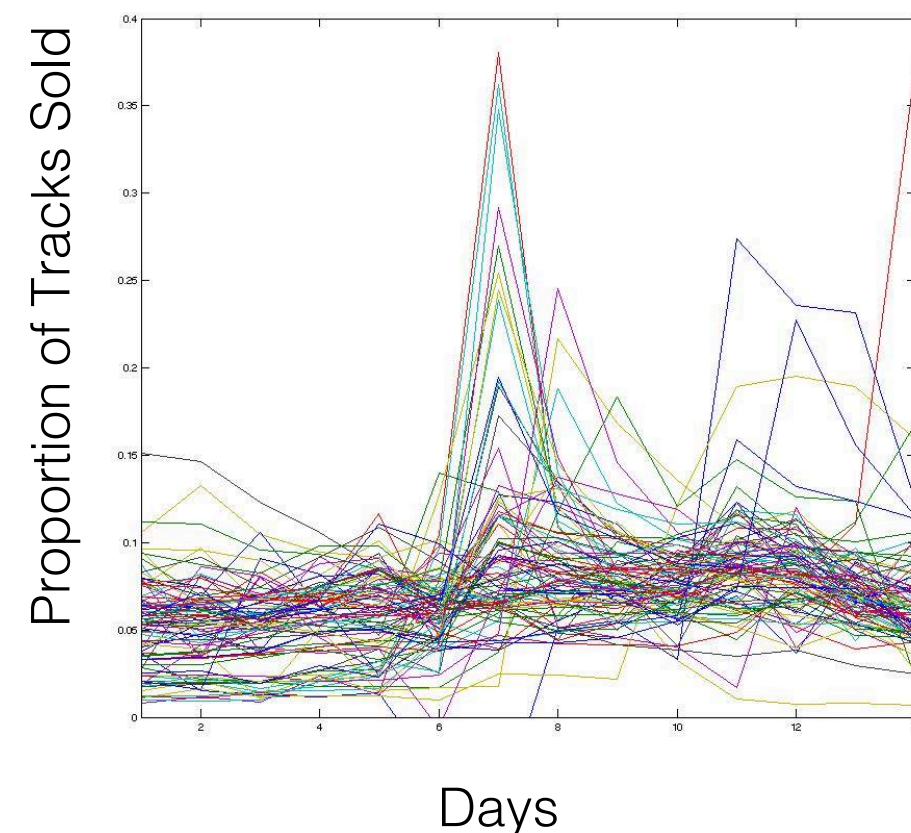
Using Social Media Metrics to Predict the Growth of Track Sales after Album Release

Pedro Garzon, Vinson Luo, Reynis Vazquez

Dataset



Evolution of Track Sales Around Album Release



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

Results

$$\text{iTunes track sale growth} = \log\left(\frac{\text{\#tracks sold 1 week after release}}{\text{\#tracks sold 1 week before release}}\right)$$

	Linear Regression	Naïve Bayes	SVR 1	SVR 2
Standard Error	0.3205	0.331	0.5521	0.2384
# Data Points	1543	1543	835	140
Features Used	Sales acceleration	Sales acceleration, LastFm, Twitter	Facebook, Twitter, Myspace	Facebook, Radio Sirius, LastFm*

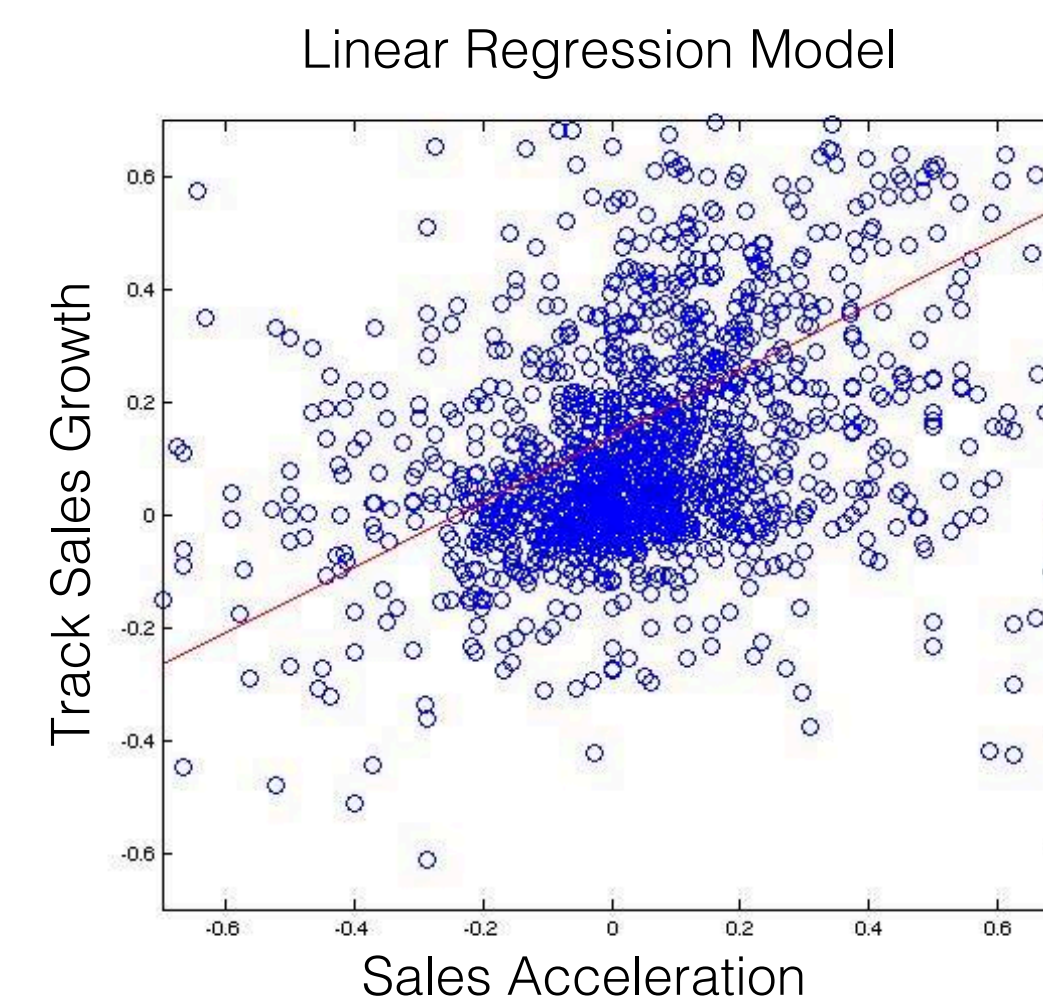
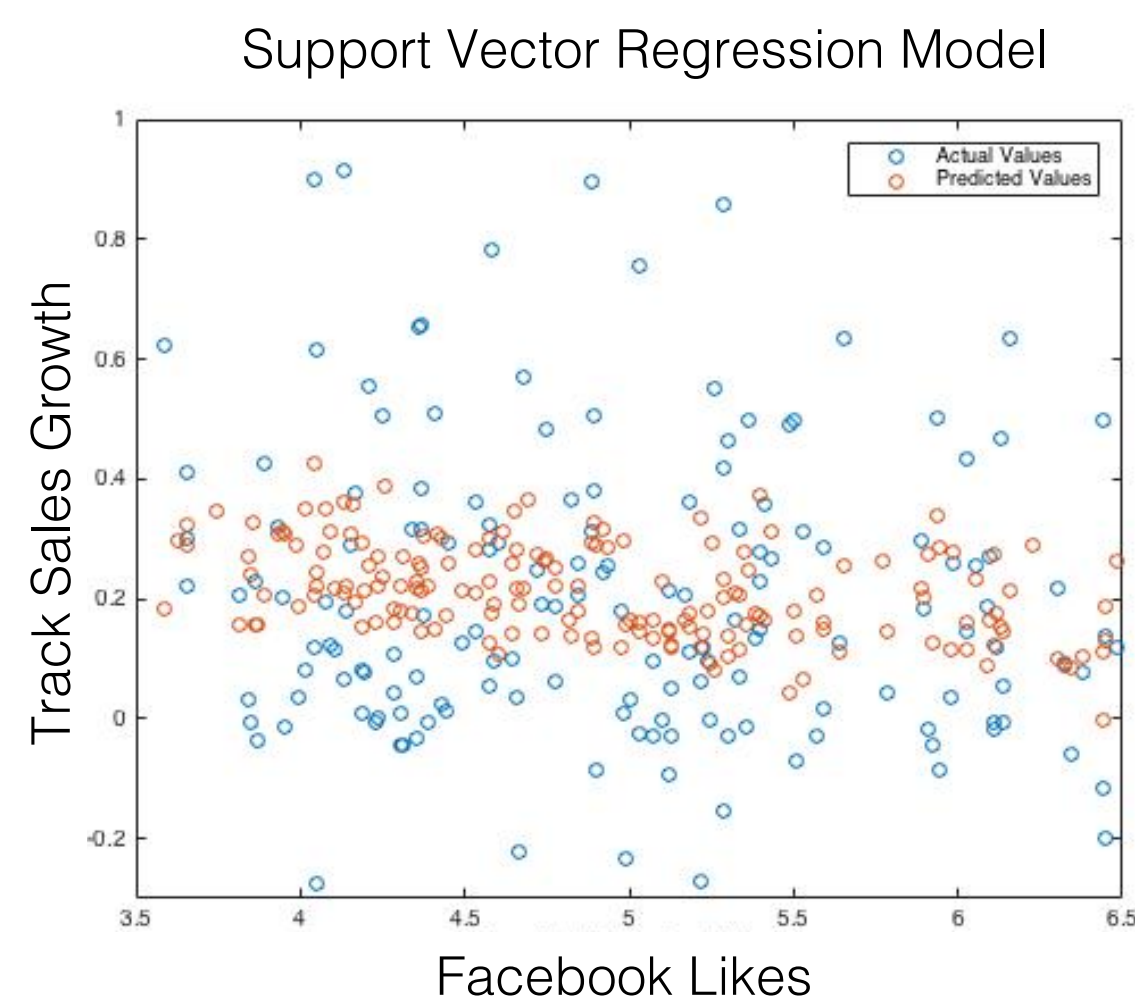
* Using K-NN Interpolation

Methods

- Processed data to look at only album releases
- Trained linear regression, naïve 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

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 naïve bayes



Challenges

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