

Readme und Einleitung in das Projekt

Projekt “EMI Music” Challenge auf Kaggle.com

von Lukas Hodel und Simon Arnold

Die Dateien in diesem Projekt

Das Projekt besteht aus folgenden Dateien:

- **README.md** - Diese Einleitung in das Projekt
- **explorative_analysis_1.ipynb** und **explorative_analysis_users.ipynb** - Erste explorative Analyse des Datensatzes bzw. speziell des **users.csv**-Datensatzes
- **predicting.ipynb** - Feature Engineering und Modelle zur Vorhersage der Bewertungen, also der eigentlichen Lösung der Aufgabenstellung
- **user_csv_plots.ipynb** und **words_csv_plots.ipynb** - Visualisierungen der Datensätze, mit Fokus auf die Datei **users.csv** und **words.csv**
- **words_cooccurences.ipynb** - Kookkurentenanalyse der zur Beschreibung der gehörten Künstler genutzten Wörter (siehe Einleitung in **bericht_arnold_hodel.pdf**)
- **bericht.bib** - Bibliographie, verwendet im Bericht.
- **bericht_arnold_hodel.md** und **bericht_arnold_hodel.pdf** - Ausgearbeitetes Paper zur Lösung der Aufgabenstellung des Projektes im Markdown- und Pdf-Format. Dies ist die finale Fassung der Projektarbeit.

Die **csv** Dateien in diesem Projekt sind die Datensätze der Challenge. Weitere Dateien dienen zur Erstellung der PDF-Datei.

Verwendete Technologien

- python 2.7.6
- ipython 3.2
- pandas 0.13.1
- numpy 1.8.2
- networkx 1.9.1
- matplotlib 1.3.1
- scikit-learn 0.16.1