Analysing music and acoustics through the lens of machine learning

Anmol Goel | Shekhar Tyagi
Chaitanya Dagar | Saurabh Rai

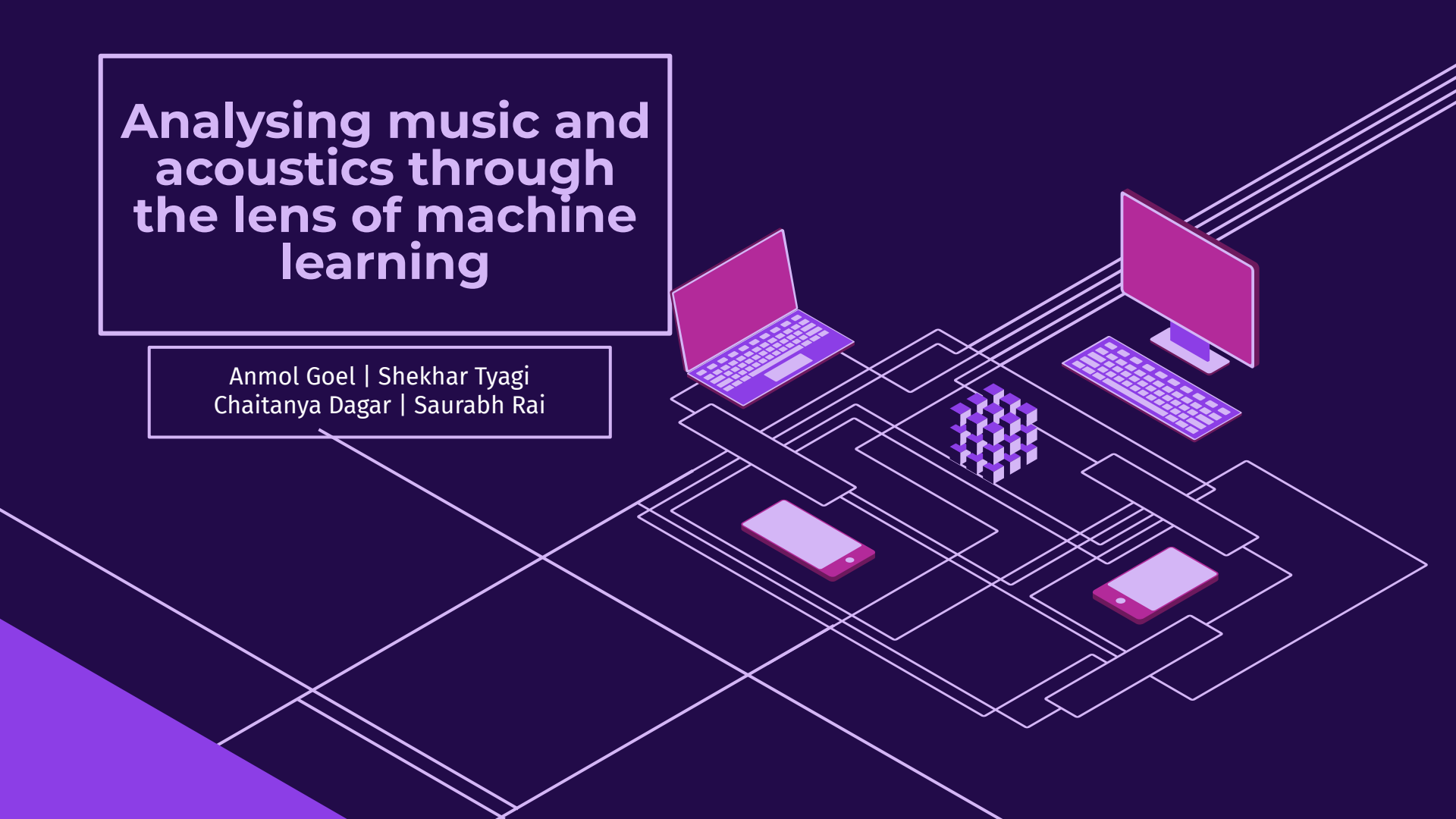


TABLE OF CONTENTS

Motivation

We analyse music and acoustic features computationally to derive insights and make recommendations

01

Introduction

PROBLEM

Music analysis is an understudied field. Lack of insights in publicly available datasets is a major problem.

SOLUTION

We aim to analyse music datasets using statistical and data visualisation tools. This will help in gaining better insights on how music can be modelled computationally.

02

Project Overview

Project Components

Recommendations

We build a robust recommender system based on acoustic features of songs.

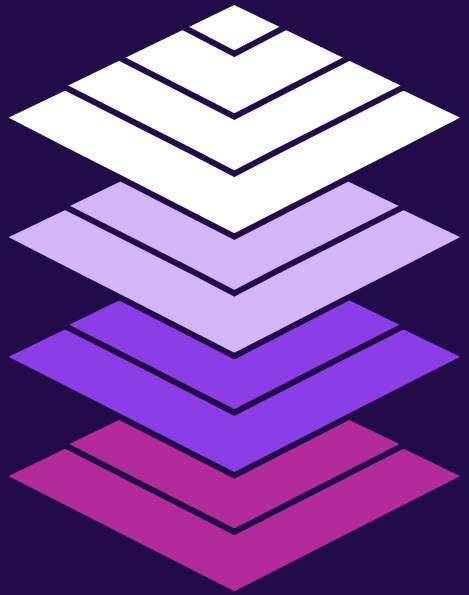
Time Series Analysis

We analyse how music has evolved through the last century and predict acoustic features based on machine learning models

Clustering

We cluster various artists and genres of music using Hierarchical Agglomerative Clustering algorithm.

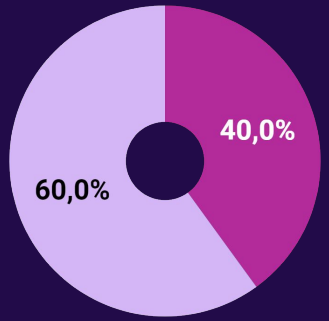
Multi Tier Framework



- **Recommendations**
  Based on cosine similarity of acoustic features.
- **Time series analysis**
  Analyse how features and song change through the years
- **Clustering**
  Cluster similar songs/genres
- **Dashboard**
  Build webapp to host the above three.

Data

Songs



1900s

2000s

Public Data

Spotify music data of songs from 1921-2020.

160000 songs and acoustic features like tempo, energy, valence, etc.

Our project is a hybrid of recommendation systems, time series analysis and natural language processing.

Acoustic Features

Popularity

How popular a song is.

Acousticness

How many acoustic instruments are in song.

Energy

How energetic the song is.

Tempo

Tempo of the song.

Danceability

How danceable the song is.

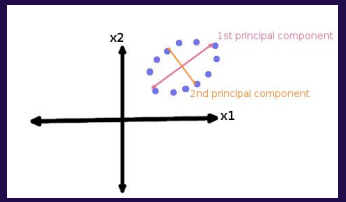
Genre

Genre of music like pop, rock, etc.

03

Algorithms

Tools & Algorithms

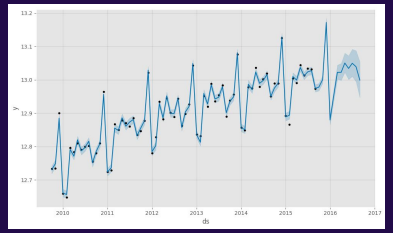


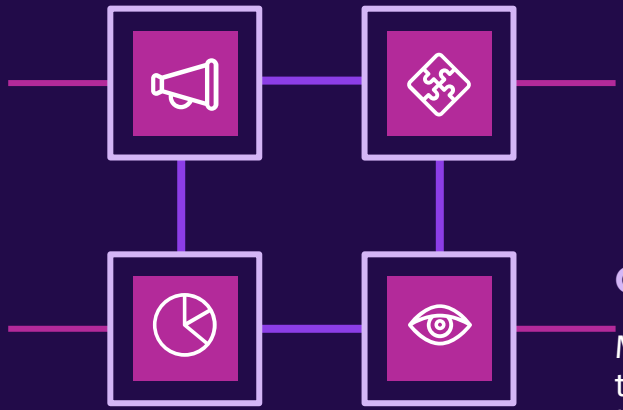
t-SNE / PCA

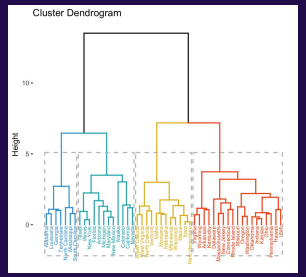
Algorithm to build low dimensional embeddings for songs using acoustic features

Prophet

Algorithm to forecast future changes and check anomalies in time series data.







HAC

State of the art clustering algorithm based on hierarchies in data.

Cosine Distance

Mathematical formula to estimate similarity between 2 latent embeddings.

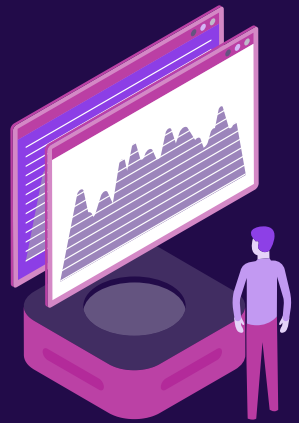
$$similarity(A,B) = \frac{A \cdot B}{\|A\| \times \|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \times \sqrt{\sum_{i=1}^{n} B_i^2}}$$

Algorithms

	Embedding similarity	Linear Time Model	HAC
Recommendation	Y	x	x
Time Series	x	Y	x
Clustering	x	x	Y

04

Real world application



Spotify

Our project can help serve better predictions and recommendations to music apps like Spotify.

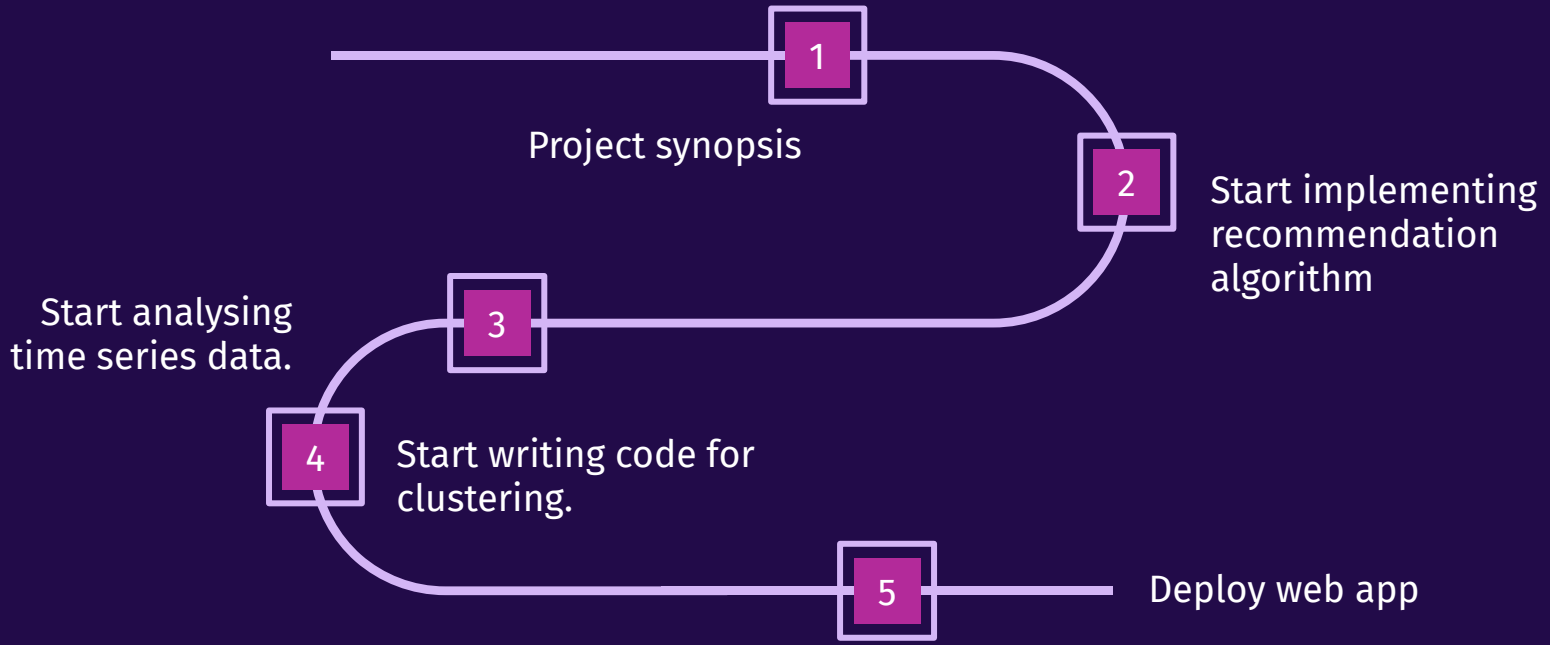
Analysis

Our project can help people to analyse music and is helpful in the field of computational social science

05

Timeline

Timeline


1
Project synopsis
2
Start implementing recommendation algorithm
3
Start analysing time series data.
4
Start writing code for clustering.
5
Deploy web app

Future Work

We aim to build a web application to visualise the musical features and recommend similar songs to users

THANKS!

Anmol | Shekhar | Chaitanya | Saurabh