Ekans - Milestone

Selin Cansu Akbaş
Student
Maltepe/Çankaya
191180005
selincansuakbas@gmail.com

Furkan Onur Akman
Student
Maltepe/Çankaya
191180009
akmanfurkanonur@gmail.com

Mert Akgüç
Student
Maltepe/Çankaya
191180006
mert.akgucc@gmail.com

ABSTRACT
In this paper, we describe what we've accomplished so far about our project, and we write what else we plan to do.

1. INTRODUCTION
K-pop has become a truly global phenomenon thanks to its distinctive blend of addictive melodies, slick choreography, and production values, and an endless parade of attractive South Korean performers. They spend years in grueling studio systems learning to sing and dance in synchronized perfection. As a music streaming and media services provider, Spotify gives millions of streamers across the globe to listen to this extraordinary music genre. According to Spotify, the topmost Top-streamed K-Pop artists on Spotify include BTS, BLACKPINK, EXO, TWICE, and Red Velvet. Hence, we will dig the insight based on these 5 idols.

2. MOTIVATION
The specific topic that we plan to work on is K-pop. K-pop (abbreviation of Korean pop) is a genre of popular music originating in South Korea. K-pop idols are groups and artists formed by the various entertainment companies creating catchy Korean popular music and targeting younger audiences. The music groups are formed from a group of people who are all particularly talented in at least one of the following: singing, rapping, and dancing. These idols often enter the entertainment company in their teens and then train hard for years in areas like singing, rapping, dancing, and foreign languages. Then, if they are lucky and talented enough, by the end of their teens they will be picked for an idol group. We chosed K-pop because it became a global phenomenon. So, we thought it would be cool to analyze K-pop using machine learning to explore interesting insights. The main questions that will be driving our project are about K-pop idol analysis. Such as idols gender or age. Also group analysis in K-pop and name analysis of this groups. We examine about their Spotify or Youtube analysis. If we talk about the exploratory questions that we plan to look for answers to, we can say that they will be as follows. How is the comparison between male and female idols number? Did the K-pop idols only come from South Korea or are there idols from other countries, nationalities as well? How many groups in K-pop? Are there any idols who have more than one group? How many idols have more than one group? K-pop idols use their original name as stage name? How many idols use stage name? How is the K-pop idols age generation? So, we find interesting insights from the variables.

We can see these relationships:

- Number of years they listened to k-pop is positively correlated with the number of hours they listen to k-pop, money they spend on merchandise, and age.
- The number of hours k-pop fans spend on watching k-pop youtube music video is positively correlated with the number of hours they listen to k-pop.
- The more time they spend on listening to k-pop, the more money they spend on purchasing k-pop merchandise.
- The more k-pop youtube videos they watch and the more k-pop they listen, the more groups they like.
- The younger they are, the more time they spend on listening to k-pop and watching k-pop videos.
- Age has nothing to do with how much money they spend on purchasing k-pop merchandise per year.

3. DATASET
Database of Kpop-Idols [1] (dbkpop) is a dedicated database of most kpop idols. the website gives important information for our questions, such as how many members are in a group, what age are the members when they debut, how many members are in a group, and how many active groups there are currently. There are also other values that may be interesting such as what country idols were born and how many idols were in a former group.

	Stage Name	Full Name	Korean Name	K. Stage Name	Date of Birth	Group	Country	Birthplace	Other Group	Gender
0	A.M	Seong Hyunwoo	성현우	에이엠	1996-12-31	Limitless	South Korea	NaN	NaN	M
1	Ace	Jang Wooyoung	장우영	에이스	1992-08-28	VAV	South Korea	NaN	NaN	M
2	Aeji	Kwon Aeji	권애지	애지	1999-10-25	Hashtag	South Korea	Daegu	NaN	F
3	Ahln	Lee Ahin	이아인	아인	1999-09-27	MOMOLAND	South Korea	Wonju	NaN	F
4	Ahra	Go Ahra	고아라	아라	2001-02-21	Favorite	South Korea	Yeosu	NaN	F
...	...	...	...	...	...	...	...	...	...	...
1305	Ziu	Park Heejun	박희준	지우	1997-06-16	VAV	South Korea	NaN	NaN	M
1306	ZN	Bae Jinye	배진예	지엔	1994-06-09	LABOUM	South Korea	Bucheon	UNI.T	F
1307	Zoa	Cho Hyewon	조혜원	조아	2005-05-31	Weeekly	South Korea	NaN	NaN	F
1308	Zuho	Bae Juho	배주호	주호	1996-07-04	SF9	South Korea	NaN	NaN	M
1309	Zuny	Kim Joomi	김주미	주니	1994-12-08	Ladies' Code	South Korea	Gwangju	NaN	F

We will also use Spotify Web API to get data. We will extract the data through Spotify's API. Spotipy is a lightweight Python library for the Spotify Web API [2]. With Spotipy you get full access to all of the music data provided by the Spotify platform. Spotipy supports all of the features of the Spotify Web API including access to all end points, and support for user authorization. Spotify has two measures of prevalence. Followers and popularity. So, we can question who are the most popular K-pop artists? Or how does popularity change? Also this K-Pop data

analysis using data provided by Spotify API which could be used to answer question like: How many K-pop albums released every year?

First, you need to register a regular Spotify account. Then, you log into the developer dashboard. There you click “create an app” and enter the basic information. Once you’re in you should see your CLIENT_ID and CLIENT_SECRET. These should be set as global variables with the same names. In this way you can ensure that you are always connected to the API. We used a Python package called Spotipy, specifically designed to work with Spotify’s API. It makes your life even easier.

Data cleaning is an important step as you want the cleanest data for EDA and model building. If you put garbage in, then you get garbage out from the model. Datasets might have leading and trailing white spaces. So, I decided to remove those white spaces using this function. Then I removed columns “Korean name” and “K. Stage Name” as it’s not useful. I assumed mostly the readers here can not read Hangul (Korean letter). I also dropped “Birthplace” because it is not a necessary feature.

4. METHOD

We will use supervised learning of machine learning techniques. It has regression, decision tree, random forest. We will be using these techniques for reading the data, cleaning the data, filtering, and moving columns or this kind of thing.

Regression analysis is a powerful statistical method that allows you to examine the relationship between two or more variables of interest. While there are many types of regression analysis, at their core they all examine the influence of one or more independent variables on a dependent variable. Regression analysis provides detailed insight that can be applied to further improve products and services.

Decision trees are excellent tools for helping you to choose between several courses of action. They provide a highly effective structure within which you can lay out options and investigate the possible outcomes of choosing those options. They also help you to form a balanced picture of the risks and rewards associated with each possible course of action.

Random forest is one of the most powerful algorithms in Machine Learning. It uses Ensemble Learning (bagging). Random forest adds additional randomness to the model while growing trees. When splitting a node, it searches for the best feature among a random subset of features instead of looking for the most important feature. Thus, it reduces the overfitting problem in decision trees and lessens the variance, improving accuracy.

Also using the Spotify API, we will take the cells of the most listened k-pop groups in recent years and graph them. We will investigate how the listening numbers of these groups have changed over the years. We analyzing continuous variables like using histogram, boxplots, and correlation matrix etc.

5. REFERENCES

[1] https://dbkpop.com/

[2] https://spotipy.readthedocs.io/en/latest/