C964 Capstone

Jacob D. Sanchez

November 8th, 2023

Task 2

**Part A**

November 8th, 2023

John Appleseed, CTO

iStream Music Inc.

1234 iStream Rd

Austin, Texas

Dear Mr. Appleseed,

The current shape of the music industry presents a challenge for iStream Music Inc., with many music albums available, the need to effectively recommend content to users is paramount. The lack of targeted recommendation system results in missed sales opportunities. The data products goal is to enhance the users experience by providing tailored music recommendations based on user profiles. This will not only increase sales, but also enhance user satisfaction with personalized content. The data product is a music recommendation system that utilizes machine learning. It will identify a user’s age and gender to predict and recommend music based on their profile. We will use a dataset that contains user’s age, gender, and music preferences. This data will be used as the base to train the machine learning model. The data will also be preprocessed, and features such as NumPy, Pandas, and Scikit-Learn will be deployed for efficient data handling and model training.

**Objectives and Hypotheses:**

Objectives: Create a music recommendation system that will help boost sales.

Hypotheses: Users that receive music recommendations based on age and gender will show a boost in sales versus without.

**Project Methodology:**

* Import the data and ensure it meets requirements.
* Clean the data.
* Separate the data into training sets.
* Create and train the model.
* Make predictions.
* Evaluate and improve.

**Funding Requirements:**

A budget of $10,000 is required for development, testing, and scalability. This includes development tools and computer hardware.

**Impact on Stakeholders:**

The execution of the music recommendation system will have a positive impact on users and boosting sales. iStream Music Inc. benefits from satisfied customers and increased profits.

**Ethical and Legal Considerations:**

There are no legal and ethical considerations as there is no sensitive data being collected, however, privacy measures will be implemented.

**Expertise Relevant to the Solution:**

I have over 5 years of machine learning experience and have deployed various applications utilizing different models to provide predictions or recommendations.

Sincerely,

Jacob D. Sanchez

Jacob D. Sanchez, Developer

**Part B**

**Executive Summary: Addressing IT Professionals**

The need for personalized music grows more and more every day. Our data product will change iStream Music Inc.’s recommendation system. This comes with a few items that align with best practices and excellent methodologies. We will address the challenge to effectively recommend music genres to users based on their age and gender, this will allow for more personalized experience. Existing customers of iStream Music Inc. are requesting a more tailored experience. The data product fulfills their needs by recommending specific music genres, enhancing user experience and increased sales. The current system lacks the ability for user profiling and personalized recommendations. Our data product will help bridge the gap by utilizing machine learning models to predict and recommend music genres based on a user’s age and gender. We are going to use existing user information that contains their age, gender, and music genre preference. No extra data collection is needed. Our methodology that we will use to guide and support the data product design and development is the utilization of a decision tree-based machine learning model. The deliverables include a verified and refined dataset, a machine learning model that has been trained, and an interface that recommends music genres based on age and gender, as well graphs that show the different age groups and genders with the recommended music genres. Our team will start with the data validation, processing and then model development and testing will be conducted. The outcome is to increase music sales and effective genre recommendation. The data will go through intensive testing, and we will compare the model’s recommendations with the ones users prefer. Users feedback will also be important to further enhance the music recommendation system. We will be utilizing Anaconda, Python, NumPy, Pandas, Matplotlib, Jupyter, and SciKit-learn for the development. Our team is proficient in machine learning models and Python, which will execute each phase for a stress-free development process. Each teammate will be assigned a task that will go directly with their experience for the human resource requirement.

**Resource Costs:**

|  |  |
| --- | --- |
| **Items** | **Cost** |
| Hardware | $2500 |
| Software Environment Costs – hosting, maintenance, and deployment | $7500 |
| Labor Time | $100 per hour |

**Projected Timeline:**

|  |  |  |  |
| --- | --- | --- | --- |
| **Milestones** | **Duration** | **Projected Start** | **Projected End** |
| Planning & Design | 80 hours | November 7th, 2023 | November 9th, 2023 |
| Development | 80 hours | November 9thh, 2023 | November 12th, 2023 |
| Documentation | 80 hours | November 12th, 2023 | November 15th, 2023 |

**Part D: Post implementation Report**

**Solution Summary:**

The proposed solution is a data product – a music recommendation system that utilizes machine learning. This system will predict genres based on users age and gender. The goal is to enhance the users experience, but ultimately increasing sales for iStream Music Inc.

This data product addresses the inaccuracies of music recommendations by using a decision tree-based machine learning model. It takes in the users age and gender to provide specific music recommendations. The development process included importing, cleaning, and training the model, making predictions and continuous evaluation for improvement. This data product aims to increase sales.

**Data Summary:**

The raw data comes from user profiles that contains the users age, gender, and music genre preference. The dataset was collected through existing user database within iStream Music Inc.

During the data product development cycle, the data underwent preprocessing using existing tools such as NumPy, Pandas, and Scikit-Learn. Data cleaning and segmentation into its training sets were also performed to ensure quality. The preprocessed data helped tremendously for the machine learning model.

**Machine Learning:**

The method that was utilized is a decision tree-based model. Decision trees are using for predictive models that can map items to conclusions about the target value. The decision tree model was developed using Scikit-Learn library in Jupyter. To train the model, we had to feed the model data to learn patterns and make predications based on age and gender. Decision trees are widely chosen for their simplicity and the ability to handle numerical data as well as categories. This model aligns with the projects goal as we can provide accurate music recommendations based on the users age and gender.

**Validation:**

Cross-validation was employed to examine the choices of the model’s performance. This helps us evaluate how well the model infers to the dataset. The validation results displayed the model’s accuracy in predicting genres based on the users age and gender. Future plans include continuous validations as the model continuous to be refined as the dataset gets updated with new data.

**Visualizations:**

A screenshot of a graph

Description automatically generated**Bar Plot**

**Bar Chart A red and blue graph

Description automatically generated**

**Line Chart**

**A graph of a person and person

Description automatically generated**

**Quick Start Guide:**

Step 1. Download the provide zip folder: ‘MusicRecommendation.zip`

Step 2. Place this folder in the Desktop or preferred directory and unzip the contents.

Step 3. Go to Anaconda: <https://www.anaconda.com/download> and then click either MacOS or Windows download.

A screenshot of a computer

Description automatically generated

Anaconda will include Jupyter and install the required libraries without having to manually install locally.

4. Proceed with the installation.

5. Open your terminal/command prompt and type ‘jupyter notebook’ this will launch Jupyter in your browser.

A screenshot of a computer

Description automatically generated

6. Once Jupyter is open, navigate to the directory where you unzipped the folder ‘MusicRecommendation’.

7. Once that is open you should see three files, Music\_Genre.ipynb, music\_recommender.joblib, and musicDataset-copy.csv.

8. Select the file ‘[Music\_Genre.ipynb](http://localhost:8888/notebooks/Desktop/MusicRecommendation/Music_Genre.ipynb)’ which should take you to the code.

9. To execute this, select the ‘Run’ play button. If you have troubles running the program, select the ‘Restart’ button then ‘Run’.

A screenshot of a computer

Description automatically generated

10. Lastly, you should see this interface:

A screenshot of a computer

Description automatically generated