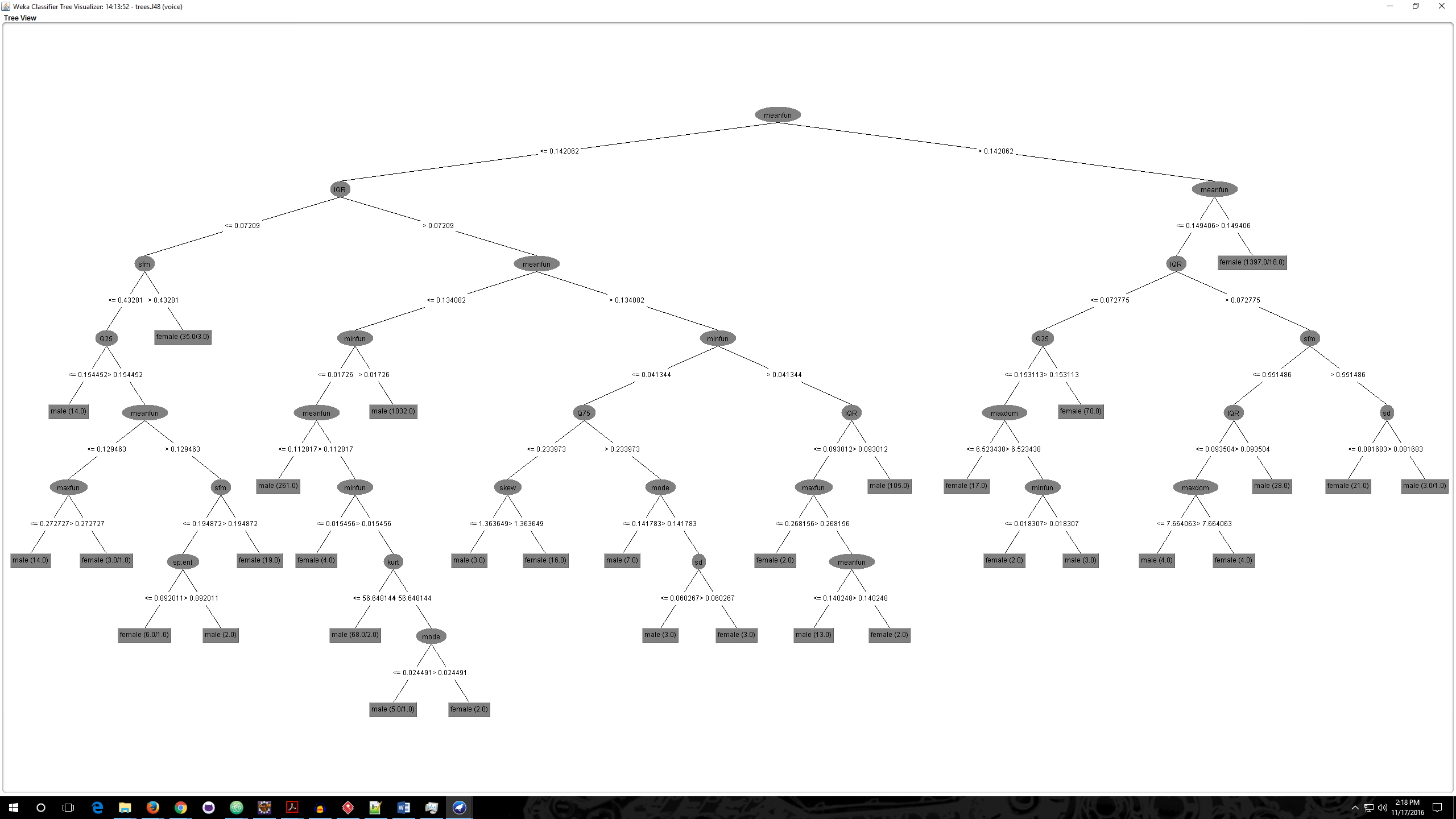
Data cleaning steps taken: We have planned to implement the recognizer in various phases which include:

1. Scrape the raw training data from the web in the form of CSV
2. The training data obtained for gender is already clean and do not contain any noisy or unwanted data.
3. Implement the Gender recognizer using python and then calculate the accuracy of the actual output obtained.
4. Research on things required to determine accent of speaker in the audio file
5. Scrape the raw data from web in the form of audio file. Calculate the required frequency from audio file, save as CSV and perform Data Integration on gender and accent training data. After integration, we would require to perform Data Cleaning.
6. The cleaned data will be given as input to the machine learning algorithm to predict gender and accent for the test audio file.

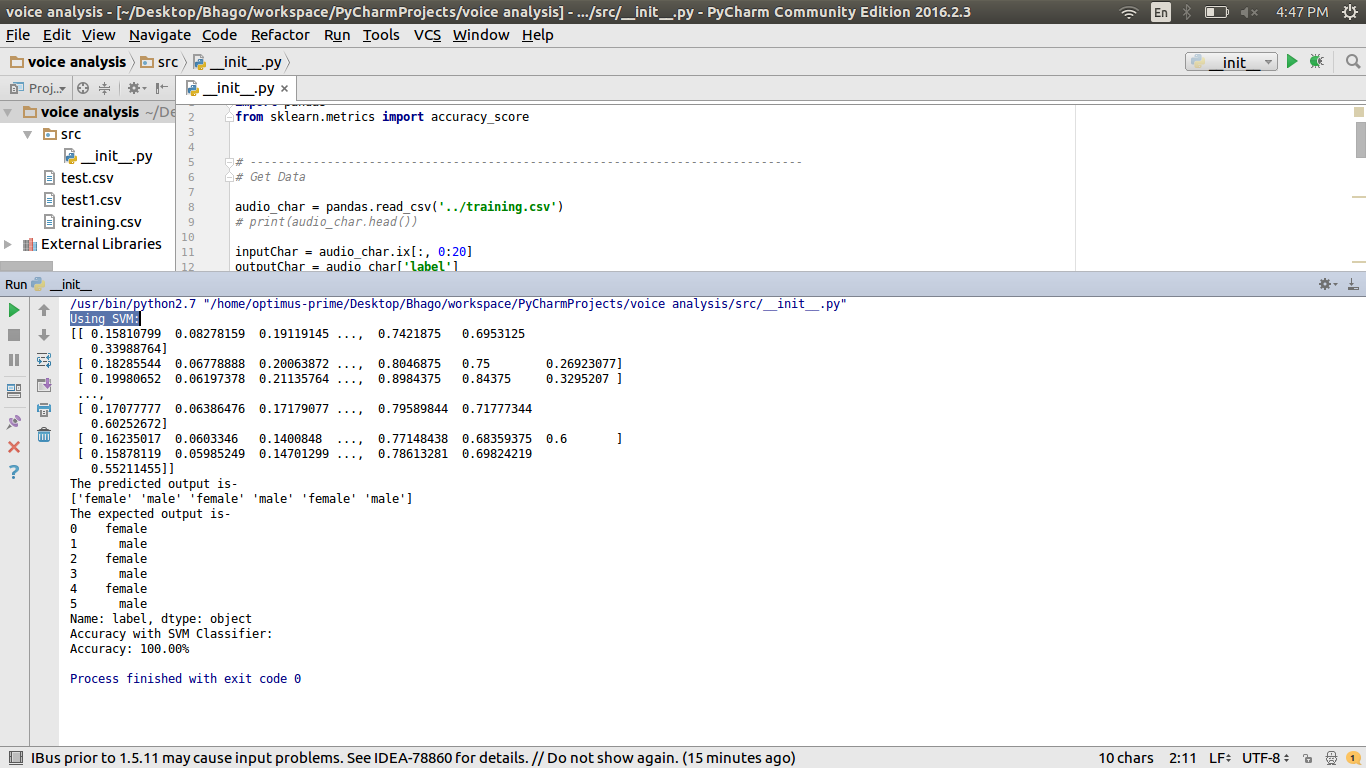
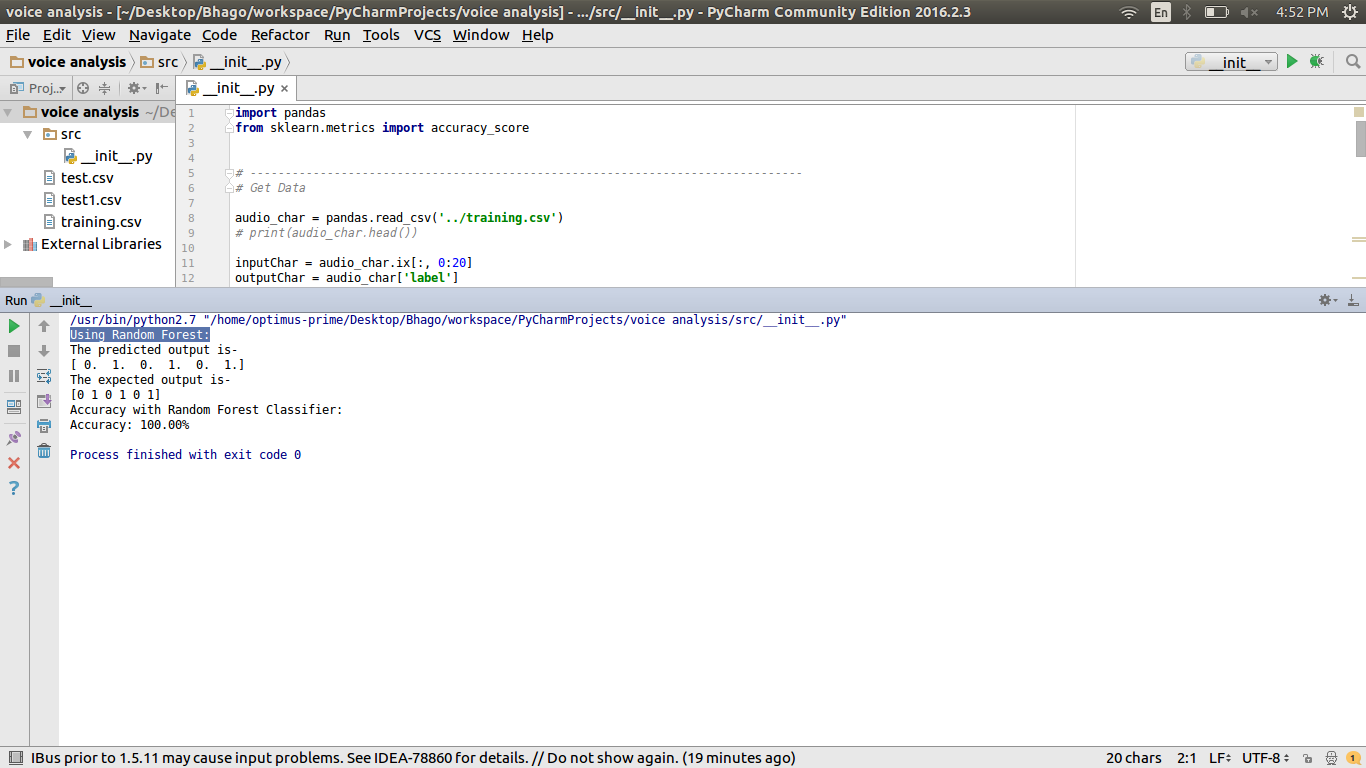
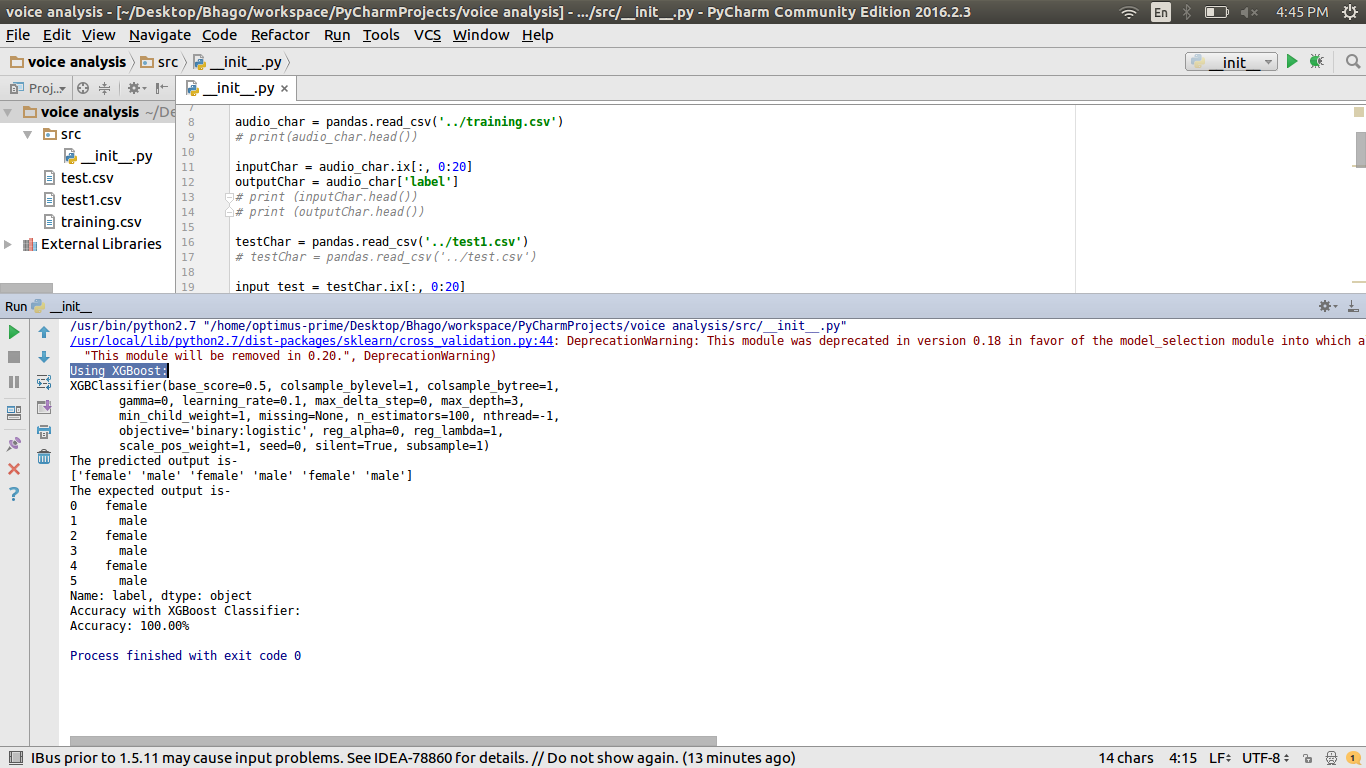
Insight:

To gain an insight, we can apply classification tree model to our training dataset to determine how these properties can be used to classify gender classification as male or female.



We can see in the above tree that the meanfun serves as a root node for detecting the gender as male or female.

Initial Screenshot:



Sketch of interaction in your final data product:

Next steps:

1. Research on things required to determine accent of speaker in the audio file
2. Scrape the raw data from web in the form of audio file. Calculate the required frequency from audio file, save as CSV and perform Data Integration on gender and accent training data. After integration, we would require to perform Data Cleaning.
3. The cleaned data will be given as input to the machine learning algorithm to predict gender and accent for the test audio file.
4. Try to find a way to determine age of the speaker in the audio