Taylor Sheneman SoftDes Mini-Project 3: Text Mining

In this project, I processed the full text of Shakespeare's *Hamlet* and the full text of Kendrick Lamar's *To Pimp a Butterfly*, the most popular rap album of 2015. Though I had not personally listened to this album in the past, I wanted to use something that is popular and that would provide a striking contrast with *Hamlet*. I Markov chained them together, in order to create something pretty weird and hopefully entertaining.

From an architecture standpoint, this was honestly kind of hacked-together. My intention was to employ a fairly straightforward architecture of *Pull separate data sources* \rightarrow *Process both into plaintext* \rightarrow *feed into Markov code*, but as it turned out, the initial scraping of the data offered plenty of challenge on its own, and programmatically processing the text would have been much more trouble than I anticipated.

The actual system architecture followed more of *Pull separate data sources* \rightarrow *Convert to editable .txt* \rightarrow *manually process files* \rightarrow *feed into Markov code*, with the Markov architecture structured like *Read through files* \rightarrow *build associative dictionaries* \rightarrow *randomly generate strings from tuple dictionary*. The most problematic section was pulling the lyrics off of Genius; it would have honestly been easier to just copy-paste all the lyrics into a text file, but I chose to learn how to use about 3 different libraries to parse through the webpages and do so automatically. I did this primarily out of stubbornness, and because I didn't mind the challenge.

The results were pretty interesting. Because the sources have such different language styles, the Markov chain tends to "chunk", or stick to the words of a single source for sentences or longer, before abruptly transitioning over into the style of the other. This is honestly pretty entertaining, at least to me, and it's interesting to see where the transitions are formed. I formatted one of the less-offensive output strings here:

Finish this, witnesses will convey just what I learned The word was respect
Just because you wore a different gang color than mine's Doesn't mean I might press the button
Had the door upon your lap?

[Sits down at Ophelia's feet.]

Oph. No, my lord.

Ham. So long? Nay then, let the runway start You know the full definition of a ring?

Oph. 'Tis brief, my lord.

Ham. Who, I?

Oph. Ay, my lord.

Exeunt Rosencrantz and Guildenstern.

How now, my lord?

Ham. Make you to ravel all this past year Pass on some advice we feel:

[Hook]

You ain't gotta lie to kick it, my nigga

[Verse 1]

I place blame on you still a fan?

[Hook 3]

The voice of the word. I took the opportunities, I worked at the end accepts his love.

Exeunt.

Oph. What is, my lord?

Ham. Nothing.

Oph. You are as foul As Vulcan's stithy. Give him heedful note; For I mine eyes into my very soul, and sweet religion makes A rhapsody of words! Heaven's face doth glow; Yea, this solidity and compound mass, With tristful visage, as against the doom,

Is thought-sick at the preacher's door

My knees gettin' weak, and my gun might blow

But we gon' be alright

[Verse 1]

This feelin' is unmatched

This feelin' is brought to you by adrenaline and good

He likewise gives a frock or livery, That aptly is put on. Refrain to-night, And that his heels may kick at heaven, And that your Grace hath screen'd and stood between Much heat and him. I'll silence me even here. Pray you be immortalised without your life

I watched you

And now you all grown up to sign this contract if that's possible

[Hook] [Outro]

I been lookin for you if you mouth it, as many of our players do, I had forgot! 'Tis so concluded on. **Ham.** There's letters seal'd; and my generation today?

[Verse 1]

I could never take the hood up, two times, deuce-four

Platinum on everythin', platinum on weddin' ring

Married to the back window, I'm a good actor.

Ham. What did you enact?

Pol. I hear him so inclin'd. Good gentlemen, give him a further edge And drive his purpose on to some confession Of his true nature, and we hate po-po

Wanna kill us dead in the hood, mob of police

Rock on the corner with a wicked tongue.

Queen.

I think the most obvious problem here is how late I'm turning this in; though to some extent this is due to my other commitments, there is something to be said for my process as well. The fact that I spent so long trying to scrape something more difficult than I had the experience to handle easily made for a significantly more painful process. I think my final results aren't bad, as I eventually got things to work and output something that is sort of interesting. If I hadn't done so, I probably would have learned something from implementing Markov analysis on my own. As is, I did at least learn some valuable stuff about different web scraping and interfacing tools, which I can certainly use in the future.