Patchalysis: A Patch Notes Analysis Tool

Sam Sui and Jovi Yoshioka

Motivation

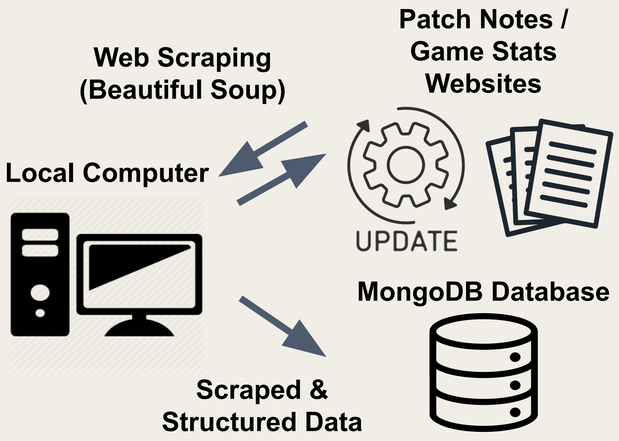
- ❏ League of Legends is an online multiplayer video-game with 160-million active monthly users.
- ❏ Beginning in 2009, the game is constantly being updated, creating an abundance of data which is difficult to analyze.

Hypothesis

Is it possible to apply natural language processing (NLP) to video-game patch notes to classify the type of change, analyzing the updates' effects on gameplay with text feature extraction and correlating it to user win rates with the modified characters?

Data Collection and Processing

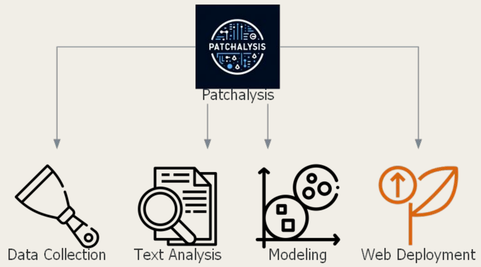
- ❏ Scrapes HTML with BeautifulSoup4 and converted the raw text into structured JSON documents.
  - ❏ **344 patch notes** since 2009 scraped.
  - ❏ **164 pro game statistics** since 2016 scraped.



Workflow

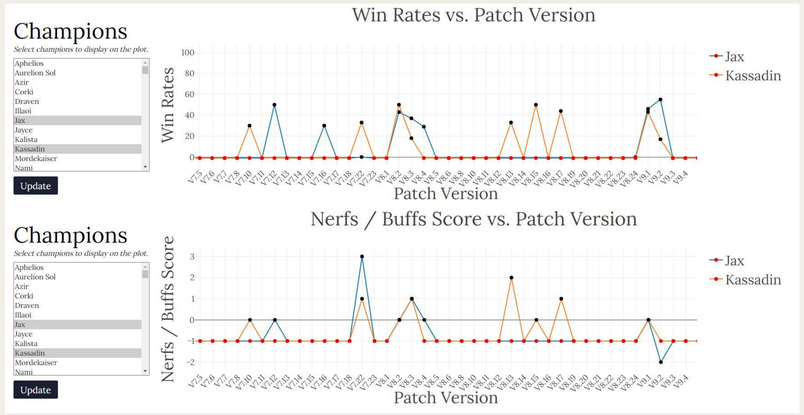
Data Pipeline

Patchalysis is multi-layered, from data collection to analysis to web deployment.



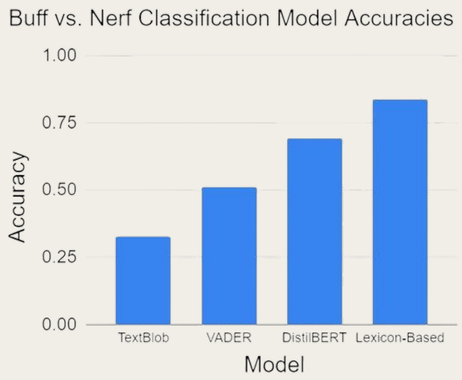
Deployed Web-Based App

An interactive app for users to visually explore patterns between character nerfs/buffs and win rates (https://patchalysis.com).



Feature Extraction and Patch Classification

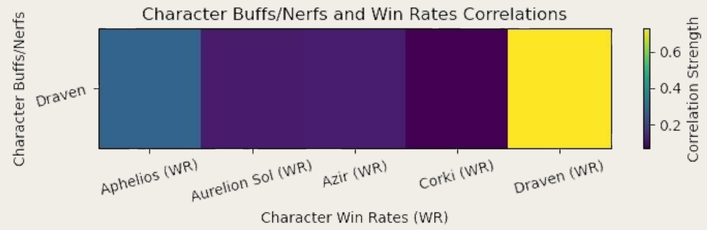
- ❏ We tested different approaches to programmatically classify an update as: a "buff" (i.e., positive, making a character stronger), "nerf" (i.e., negative, making a character weaker), or neutral.
  - ❏ The pre-trained models performed worse than the lexicon-based approach.
  - ❏ The lexicon-based approach defines keywords that, if in text, deterministically classify that text as the sentiment of those keywords.



Results

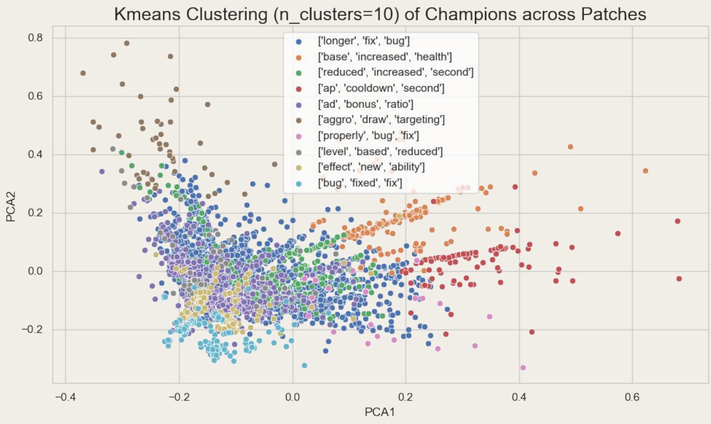
Correlation Modeling

- ❏ Uses pandas to create a correlation matrix of buffs / nerfs to win rates.
- ❏ Visualizes how a character's buffs / nerfs might affect (i.e., correlate to) other characters' win rates (e.g., as seen in the below figure).



TFIDF K-Means Clustering

- ❏ Clusters **TFIDF vectorizations** with K-Means and labels.
- ❏ Uses **NLTK** to process the text and **PCA** to reduce to two dimensions.
- ❏ Shows samples of keywords being associated with champions over time.



Conclusion & Future Work

- ❏ As demonstrated by the **84% classification accuracy**, it is concluded that NLP can successfully analyze patch notes.
- ❏ A **relationship between buffs / nerfs and win rates** was shown to exist via correlation modeling and TFIDF-based K-means clustering.
- ❏ In the future, we hope to leverage a **deeper model** and a **higher volume of text data** to better navigate game terminology.

TICKLE COLLEGE OF ENGINEERING
THE UNIVERSITY OF TENNESSEE, KNOXVILLE

This project was a part of the COSC 426: Introduction to Data Mining & Analytics course taught at the University of Tennessee Knoxville. The course is taught by Dr. Michela Taufer, accompanied by teaching assistants Ian Lumsden, Nigel Tan, Paula Olaya, Jack Marquez, and Syed Meerza.

Logo Generated By: OpenAI. (2023). ChatGPT (April 2024 version) [Large language model]. https://chat.openai.com/chat