**VraTeloka Theme Park Review**

**VraTeloka** is an application that provides review for theme parks, they want to make a new feature for their application to judge if a review is a positive or a negative one using **Natural Language Processing**. You as their programmer are tasked to make a program as a prototype for this new feature.

* When the **application is** **launched**, it will **check** if there is a **“model.pickle”** file in the **directory** or not.
* If the **file exists**, then the **application will read and load** the model from the file.

Text

Description automatically generated with medium confidence

***Figure 1. model.pickle successfully loaded***

* If the **file doesn’t exist**, the program will **train a model** from the **data provided** using **Naïve Bayes Classifier** from **NLTK**. The training data will follow **these rules**:
  + **Preprocess** the dataset by **tokenizing** the words, **remove stopwords**, **remove symbols** and **number**, **stemming**, and **lemmatizing** the words.
  + Split the data **30%** as the **testing data** and **70%** as the **training data**.
  + Compare the **review content**, and the **rating given**. If the **rating is 3 or higher**, then set the review as a **positive review**, **otherwise** set the review as a **negative review**.
  + **Train** the model using **Naïve Bayes**.
  + Show the **3 most informative features** and the **training accuracy**.
  + **Save training model** to pickle file with the name **“model.pickle”**.

Text, letter

Description automatically generated

***Figure 2. Most informative features and training accuracy***