

Project Proposal

-Problem statement:

Many parents suffer in observing what their children read, as they fear that they will read stories that do not suit their ages, such as horror stories. And also the owners of websites dedicated to story books suffer in classifying the appropriate stories for children, so we decided to make a model capable of classifying the appropriate and the inappropriate stories for children.

-Data Description:

We do web scraping from Arabic site (Qssas) specialized in publishing various and renewed stories. We decided to focus on children's stories, horror, police, and wars. The dataset consists of 1020 Rows and 1 feature, which is a text.

-Tools:

- Programs: Jupyter Notebook
- Libraries: numpy , pandas , matplotlib , seaborn , csv, nltk , requests , BeautifulSoup , re , string , LogisticRegression , sklearn
- Functions: shape , value_counts , strip , replace , info , drop , astype , groupby , mean , get , content , find_all , zip_longest , writer , writerow , fit , predict , OLS , len , where , train_test_split , CountVectorizer , confusion_matrix , round , MultinomialNB , BernoulliNB , concat , TfidfVectorizer ,
- Plots: pointplot , barplot , histogram , scatter , distplot , Pie Chart.

-MVP Goal:

Create a model that is able to classify the type of story if it is suitable for children or not with a high accuracy using the clustering model