**Project –Offensive language detection model in the adult industry**

Team members: Cîrstea Ariadna Alexandra, Liviu Nichifor, Adrian Gigi Postolache

Link GitHub: [https://github.com/AriadnaAgnis/CommentsAnalysis](https://github.com/AriadnaAgnis/CommentsAnalysis?fbclid=IwAR1VHz6n0M-QJTiMOIbZR0vQxxzhIxI-ybpiNsEqxlaq7egogdDfgSQwM_Y)

**Project’s short description** *(max 50 words): Our project wants to come with the best solution to detect offensive language in comments to prevent online bulling. PornHub is one of most vulnerable platforms for this type of aggression because of the nature of the content.*

**Architecture (schema):**

Data modelor Regression

Python cleaning algorithm

Comment scoring/ Filter system

Web scraper

**Project’s details** *(max 1000 words), including:*

1. Motivation: One of the largest industries on the planet has to potential to influence human behavior. Implementing a filter for online aggression is mandatory to protect the users.
2. *Challenges: In the adult industry words that might be consider offensive in other platforms have a different connation. Normal technical approach is not effective and we need to develop a customized solution.*
3. *State of the art: Python sentiment analysis, Regression classification*
4. Short description of each module in the architecture, including methods to be used, programming language, additional resources (if any), input/output formats, etc.

* Data set training- Data set with classified language
* Data set scoring- All comments that are gathered via web scraping
* Web Scraper- technical solution to gather comments from web pages
* Python cleaning algorithm- solution for noise reduction of the gathered data
* Data modeler regression- Statistical logic to determine classification of the data set (true/false)
* Comment scoring filter system- Final output

1. Description of the integration of different modules: problems foreseen, if any (incompatibility of programming languages, input/output format changes, etc.).

Gathered data was done manually, Lemmatization proves ineffective because of different spelling errors and signs

**Project management** (no. max. words no. imposed):

1. Assignment of team members for each different module (one team member per module)

* Ariadna Alexandra – Data gathering, Lemmatization process
* Liviu- Regression solution and data modeling
* Adrian Gigi- Project management, Documentation, Landscape technical analysis

1. Will the project have a team leader? If yes, who is it? What are its main duties?

* Ariadna Alexandra- team leader. Main duties to keep track of progress and coordinate activities

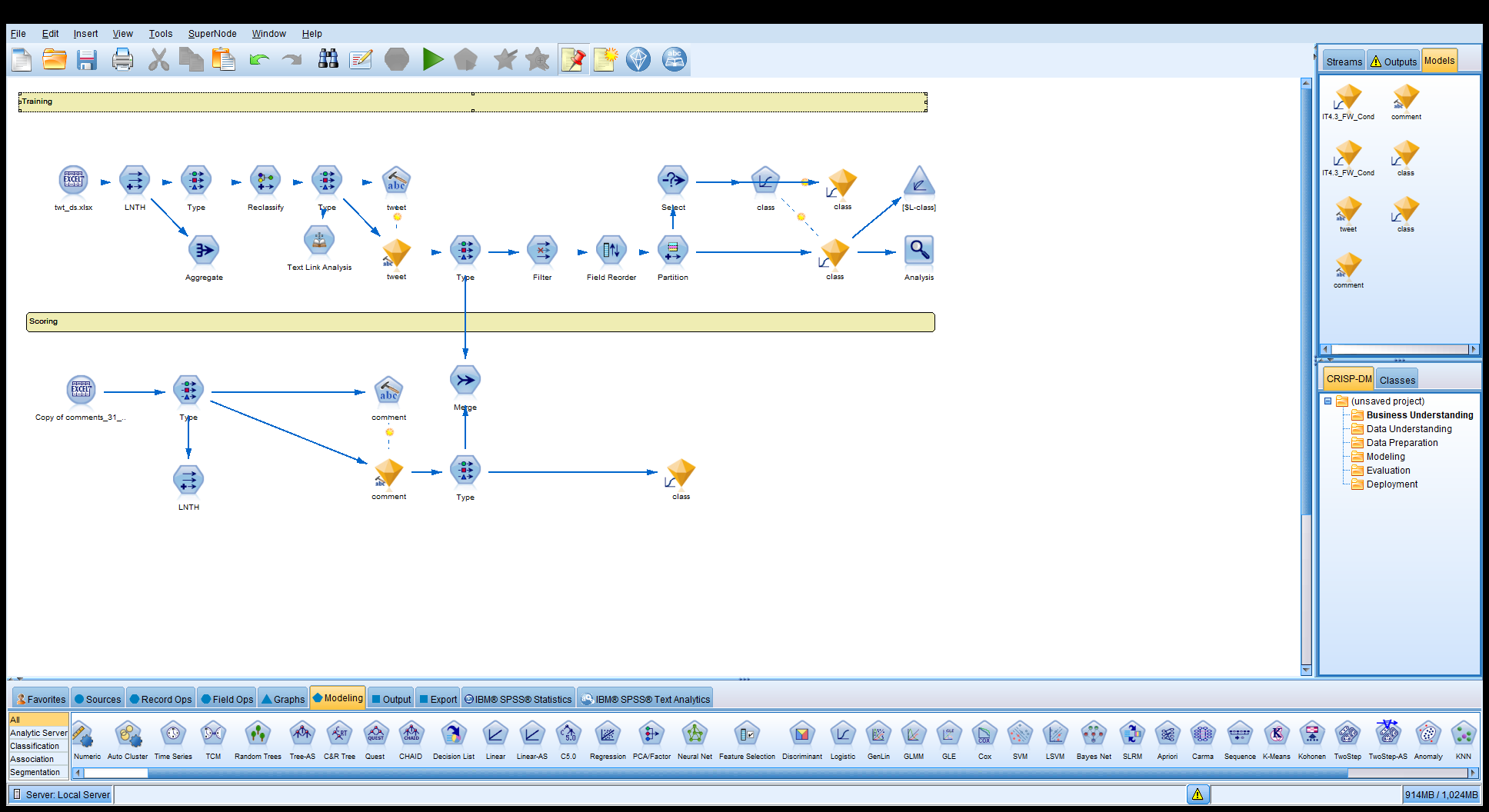
1. Detailed deadlines and milestones (for each subtask of each module, but also for the inclusion of all modules in the final system). Deadlines are actual dates.

* Milestone- Lemmatization process is difficult, deadline 18,01,2020

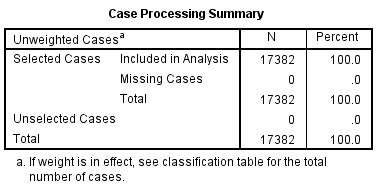
1. Qualitative measures for the evaluation of the project (e.g. no. of annotations, no. of sentences processed, etc.). These measures **will** be used to assess the projects.

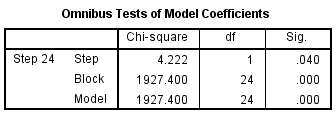
**Final model – Logistic Regression – Forward Stepwise LR**

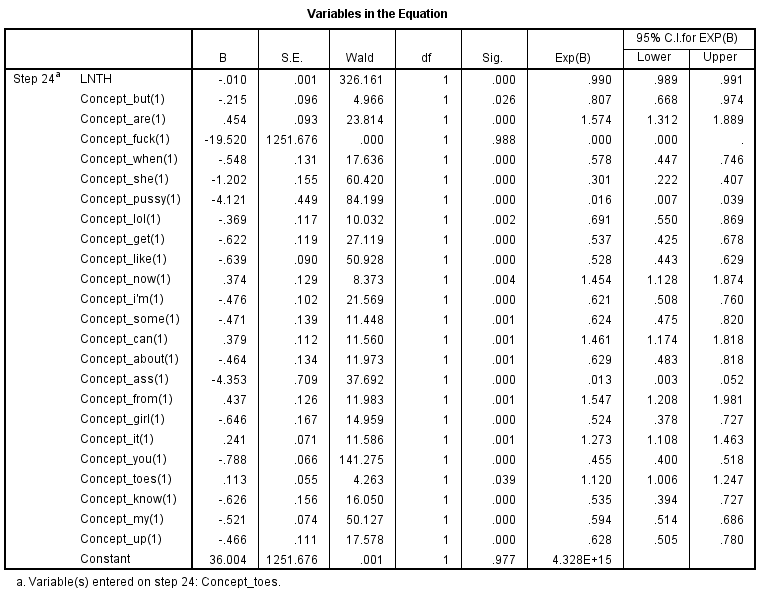
1. Training and testing the model
   1. For training/testing we will use Twitter dataset where comments are classified;
   2. Prior to that we will bring PornHub comments and Twitter comments into the same standard format that can be used for regression analysis;
   3. The model will be generated on Twitter training sample (70%) and tested on the remaining 30%;
   4. At the end we will apply the model on PornHub scoring dataset and classify each comment;
2. Model architecture

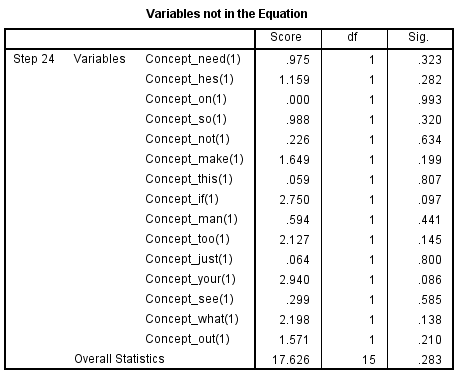


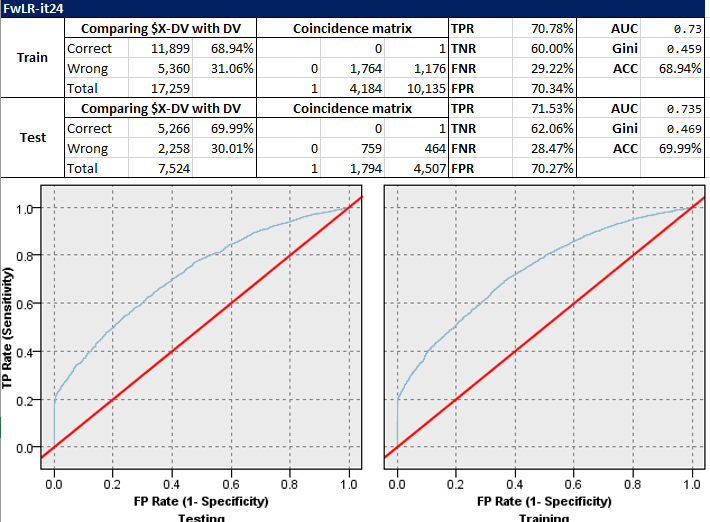
1. Model measurement











The FW Stepwise method created 24 iteration, the last one being the best one containing only significant variables. The overall accuracy is ~70% which is decent. Performance between Training and testing is similar, so it is a robust model which can be used for scoring.