**Automated Workflow Solution**

Introduction

My automated workflow solution focuses on social media analysis and sentiment analysis of tweets, but its capability could be expanded in the future to cater for a wide range of media coverage and ML analysis of various metrics. For example, manually checking the validity of datasets, such as whether spokespeople are mentioned in the coverage or whether there is an anomaly in the reach figures for a particular publication, is something that could easily be done by an ML process. My product will be almost entirely automatic and will only require minor inputs from an analyst, mainly by providing a light-touch quality check of certain deliverables.

Here is the outline of the project:

Diagram, schematic

Description automatically generated

# Appendix D – Code for Data Product

STEP 1: Google Colab – Fine-tuning and evaluating the Transformers Model

Graphical user interface, text, application, email

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

Text

Description automatically generated

This output will allow for a proper evaluation of the model performance – providing tweets with both predicted and actual sentiments followed by the f1 score for each sentiment class.

Once the model has been evaluated in this way it can be used on the larger datasets to populate client reports.

The next part is to create the data pipeline that will clean and apply sentiment analysis to the tweets.

**Diagram, schematic

Description automatically generated**

Data\_pipeline.py

Text

Description automatically generated

Text

Description automatically generated

Data\_cleaning\_for\_ML\_app.py

Graphical user interface, text, application

Description automatically generated

Data\_cleaning\_for\_analysis\_app.py

Text

Description automatically generated

Data\_cleaning\_tools.py

Text

Description automatically generated

Text

Description automatically generated

Graphical user interface, text, application

Description automatically generatedRequirements.txt file

Conclusion

My project has been successful in satisfying the aims that I identified at the outset. The performance of the machine learning model has been crucial because this technology will ultimately provide much of the benefits in terms of increased efficiency and productivity that is one of my core aims. I am happy that the testing of the model revealed a 90% accuracy rate when the model was trained with 1,000 rows of labelled data and would mean that potentially it could manage the millions of tweets that would require analysing were Onclusive to offer this service across their client base.