Proceedings of the Fifteenth International AAAI Conference on Web and Social Media (ICWSM 2021)

A Dataset of Multidimensional and Multilingual Social Opinions for Malta's Annual Government Budget

Keith Cortis and Brian Davis
ADAPT Centre, Dublin City University, Glasnevin, Dublin 9, Ireland
keith.cortis@adaptcentre.ie, brian.davis@adaptcentre.ie

Abstract

This paper presents three high quality social opinion datasets in the socio-economic domain, specifically Malta's annual Government Budgets of 2018, 2019 and 2020. They contain over 6,000 online posts of user-generated content in English and/or Maltese, gathered from newswires and social networking services. These have been annotated for multiple opinion dimensions, namely subjectivity, sentiment polarity, emotion, sarcasm and irony, and in terms of negation, topic and language. These datasets are a valuable resource for developing Opinion Mining tools and Language Technologies, and can be used as a baseline for assessing the state-of-the-art and for developing new advanced analytical methods for Opinion Mining. Moreover, they can be used for policy formulation, policy-making, decision-making and decision-taking. This research can also support similar initiatives in other countries, studies in the socio-economic domain and applied in other areas, such as Politics, Finance, Marketing, Advertising, Sales and Education.

1 Introduction

Opinion Mining (OM), also referred to as Sentiment Analysis, is a popular[1] and extremely valuable research area, especially for the exploitation of user-generated content extracted from social sources, such as social media platforms and newswires commenting sections (social data). OM is considered a challenging Natural Language Processing (NLP) problem, especially when applied on social data. This evolving field, also known as Social Opinion Mining (SOM) (Cortis and Davis 2020), is tasked with the identification of several opinion dimensions, such as subjectivity, sentiment polarity, and emotion, from user-generated social data represented in various media formats that is spread across heterogeneous sources. There is a great need for quality datasets in this field, especially multilingual datasets, multidimensional opinion datasets, and multilingual multidimensional opinion datasets that go beyond Sentiment Analysis.

In this paper we present three high quality datasets focusing on bilingual multidimensional OM for the Maltese less-resourced language and English, in the socio-economic domain, specifically Malta's annual Government Budget. During this annual event, the Government presents an estimate of its expenditures and revenues for the upcoming year. These datasets cover the 2018[2] , 2019[3] and 2020[4] budgets and consist of over 6,000 online posts from newswires and social networking services. To our knowledge, this is the only user-generated content Government Budget dataset that is available for OM.

This multidimensional OM approach handles the identification of five different social opinion dimensions, namely:

- **subjectivity**: determines whether a sentence expresses an opinion– in terms of personal feelings or beliefs– or factual information– objectivity (Liu 2010);
- **sentiment polarity**: determines the polarity (positive/negative/neutral) of an expressed opinion (Liu 2010);
- **emotion**: refers to a person's subjective feelings and thoughts, such as love, joy, surprise, anger, sadness and fear (Liu 2010);
- **sarcasm**: holds the "characteristic" of meaning the opposite of what you say, but unlike irony, it is used to hurt the other person; and
- **irony**: used to convey the opposite meaning of the actual things you say, but its purpose is not intended to hurt the other person.

Malta's two official languages, Maltese (Malti), a Semitic language written in the Latin script which is also the national language, and English, were chosen. These align with Malta's Strategy and Vision for Artificial Intelligence (Schembri 2019), where the country is investing in the development of Maltese language resources and tools. This work shall also counter the threat of "digital extinction" for the Maltese language, which has low technological support available in comparison with other European languages (Rosner et al. 2012).

A recent study by (Cortis and Davis 2020) identified several research gaps from the existing literature within the SOM research area. In fact, the three datasets presented here

[1]See Google Trends interest analysis: https://trends.google.com/trends/explore?q=sentiment%20analysis

[2]https://www.finance.gov.mt//en/the-budget/pages/the-budget-2018.aspx

[3]https://www.finance.gov.mt/en/The-Budget/Pages/The-Budget-2019-G5J3D1.aspx

[4]https://www.finance.gov.mt/en/The-Budget/Pages/The-Budget-2020-GD-9691.aspx

focus on the following ones:

- Gathering of social data from more than one data source, namely social networking services and newswires;
- Enabling multiple techniques to be explored for classification purposes;
- Collection of a new social dataset for a real-world application area which contains bilingual data (Maltese and English); and
- Annotations of **five** social opinion dimensions within the data (subjectivity, sentiment polarity, emotion, sarcasm and irony).

These datasets will encourage other researchers working on low-resourced languages to create similar datasets to safeguard their languages from a technological standpoint.

A first version of the social dataset was published in 2019 (Cortis and Davis 2019), which consists of social opinions for the Malta Budget 2018. This volume has been increased with the user-generated data of the 2018, 2019 and 2020 budgets and further enhanced with the social opinion dimensions previously mentioned and the annotation types defined in the sub-sections below.

2 Impact

Governments are increasingly using political and socio-economic online user-generated content created across social media platforms and other websites to get a better grasp of the citizens' perceptions and needs, and society's problems at large. This resulted in the development of several information and communication tools and technologies, with the most critical ones being in the OM area (Charalabidis, Maragoudakis, and Loukis 2015). According to Eurostat (2019 statistics), 13% of individuals living in Malta post opinions on civic or political issues via websites, such as blogs and social networks. Consequently, these datasets of user-generated content provide a voice to the citizens who use social media platforms to make their opinions known and/or provide feedback about any particular measure announced by the Government, whether it is tax related, industry specific, or any other social initiative.

In terms of impact, the annotated datasets of social opinions for the Malta Government Budget have the potential of being used for initiatives by the Government to capture the public opinion and perception. These valuable insights can be evaluated and taken in consideration for revision of measures and/or any bills presented and discussed in Parliament. In fact, recently Malta's Minister for Justice emphasised on the importance of personal opinions that are expressed on social media and blogs, and mentioned that these are a very important source of information for the Government when carrying out certain initiatives and processes, such as the rule-of-law reforms (Xuereb 2020).

3 Process

This section describes the multi-stage process used for building each of the three natural language Malta Government Budget datasets (2018, 2019, 2020), namely, the methods employed for data collection, the annotation process of each dataset, and the data quality measures carried out to consolidate the final datasets.

3.1 Data Collection

The datasets were collected from the following data sources: **Newswires** - Times of Malta[5], MaltaToday[6], The Malta Independent[7]; and **Social networking services** - Twitter[8].

Similar to (Cortis and Davis 2019), the data source selection was based on citizens' preference for online news in Malta, with the Times of Malta and MaltaToday being the top two, followed by The Malta Independent in fourth (Martin 2020). On the other hand, Twitter is actively used for Maltese politics, especially during each annual Government Budget. This is reflected in (Kemp 2019), where the total advertising audience on Twitter in Malta amounts to over 60,000 monthly active users.

Newswires Table 1 presents the following newswires' information for each respective budget **(row 1)**: initial number of comments collected from Times of Malta **(row 2)**, MaltaToday **(row 4)**, The Malta Independent **(row 6)**, and the total number of comments left for each newswire after removing images or the ones deleted by the editor/comment owner **(row 3, row 5, row 7)**.

Newswire comments	**Budget 2018**	**Budget 2019**	**Budget 2020**
Times of Malta (Initial)	253	354	275
Times of Malta (Total)	**249**	**350**	**270**
MaltaToday (Initial)	178	296	349
MaltaToday (Total)	**175**	**280**	**306**
The Malta Independent (Initial)	46	10	39
The Malta Independent (Total)	**45**	**9**	**39**
Overall (Total)	**469**	**639**	**615**

Table 1: Details of Newswires data for each dataset

The online news articles selected from each newswire for each budget year contained content in one of the following categories:

1. overview of the upcoming budget, published either on the day prior to the budget announcement or on the day of the budget, a few hours before the announcement;
2. near to real-time live updates of the budget measures being presented for the upcoming year; and
3. overview or feedback on the presented budget, published after the budget finishes, on the same day or the following day.

These articles enable citizens to post their opinions and/or reactions on the budget and the content published in the said articles. Therefore, the articles that produced most comments, in terms of volume from citizens, were selected from each newswire. It is important to note that the majority of

[5] https://www.timesofmalta.com/
[6] https://www.maltatoday.com.mt/
[7] http://www.independent.com.mt/
[8] https://www.twitter.com/

these user comments are similar in nature to online posts published on social networking services, such as Facebook.

All of these comments were manually extracted for our datasets, in order to annotate them in terms of the different social opinion dimensions mentioned in Section 1. Moreover, four online articles from each newswire were chosen for each budget. This ensured a diverse sample of online posts[9] that reflect the opinion of the general public with respect to the budget as a whole. Therefore, budget domain specific articles, e.g., an article focusing only on Technology budget measures, were omitted from the ones selected, with priority given to the ones reviewing the budget at large and ones that listed or gave an overview of all the budget measures for each domain.

A total of 1,800 comments were collected from the selected newswires for the 2018, 2019, and 2020 budgets. The ones that resulted in deleted comments (by the respective newswire or comment owner) or comments that consisted of images only were removed, leaving a total of 1,723 online posts for annotation purposes.

Social networking services As for online posts from Twitter (tweets), the ones that contained the following hashtags and/or keywords were extracted for each of the three budgets: "maltabudgetYY", "malta budget YYYY", "maltabudget YYYY", " malta budgetYYYY", "maltabudgetYYYY", "malta YYYY budget", and "YYYY budget malta", with "YY"/"YYYY" referring to the respective budget year "18"/"2018", "19"/"2019", "20"/"2020". The chosen keywords were based on the manual identification of the most common keywords used in the content of tweets relevant to the Malta Budget.

Table 2 presents the following Twitter data information for each respective budget **(row 1)**: date range of the data collection period for each budget where the date of the first and last tweet were determined through a manual search using the Twitter Advanced Search feature **(row 2)**, the total amount of tweets initially collected using the seven hashtags and/or keywords **(row 3)**, the total amount of tweets remaining after removing duplicate (based on exact content) tweets and retweets **(row 4)**, and the official budget hashtag used by the Government of Malta.

Budget	**2018**	**2019**	**2020**
Data collection dates	28/08/2017-05/06/2018	20/07/2018-23/04/2019	01/09/2019-25/03/2020
Tweets collected (Initial)	4,168	4,682	4,904
Tweets remaining (Total)	**1,673**	**1,677**	**1,314**
Offical budget Hashtag	maltabudget18	maltabudget19	maltabudget20

Table 2: Details of Twitter data for each dataset

[9]"Online posts" is the general term used within this paper to refer for both comments and tweets

A total of 13,754 tweets were collected from Twitter for the 2018, 2019, and 2020 budgets. Any duplicate tweets and retweets (based on exact content) were removed, leaving a total of 4,664 tweets for annotation purposes.

The Twitter Premium Search API[10] was used via the TwitterAPI Python library[11] (used to access the Twitter API) to collect the tweets related to the three budgets, in particular the full-archive data endpoint. No online posts from Facebook were collected given that access to the Public Feed API[12] is restricted and users cannot apply for it.

Sampling Strategy A random sampling strategy was used to gather the data. Four online articles for each newswire were chosen (in total twelve articles), specifically ones that had the highest number of user-generated comments. As for Twitter, all data made available within the limits of the respective Twitter API was gathered, therefore equating a significant representative sample of the population.

3.2 Annotation

All the online posts collected from the newswires and social networking services were presented to three raters. In terms of expertise, all of the raters were proficient in Malta's two official languages (Maltese and English), with two raters being computer science graduates and working in the technology domain and one rater being a business and management graduate and working in the human resources domain.

All the raters were given a lecture on OM, whereas annotation guidelines[13] were provided to support them during the annotation process. These guidelines were piloted twice during the annotation process –primarily after twenty-five (25) annotations, and then after a hundred (100) annotations–, following clarification and feedback with the raters. Each rater took approximately 120 hours to annotate the three datasets, therefore the total estimated annotator time is 360 hours. A fourth rater, a computational linguist from academia, consolidated the annotation values to create the three final datasets. The annotation process discussed above follows the Model (Model and Guidelines) - Annotate - Model (Evaluate) - Annotate (Revise) cycle defined in (Pustejovsky and Stubbs 2012).

Each online post is annotated with the following information (annotation types):

1. **Subjectivity**: binary value, with 1 referring to subjective posts and 0 referring to objective posts;
2. **Sentiment Polarity**: categorical value (3-levels) for the sentiment polarity of the online post (negative, neutral, positive);
3. **Emotion**: categorical value for the emotion of the online post based on Plutchik's (Plutchik 1980) eight primary emotions (joy, sadness, fear, anger, anticipation, surprise, disgust, trust);
4. **Sarcasm**: binary value, with 1 referring to sarcasm in online posts;

[10]https://developer.twitter.com/en/docs/twitter-api/premium/search-api/api-reference/premium-search

[11]https://www.github.com/geduldig/TwitterAPI

[12]https://developers.facebook.com/docs/public_feed/

[13]https://www.git.io/JOqss

5. **Irony**: binary value, with 1 referring to irony in online posts;
6. **Negation**: binary value, with 1 referring to negated online posts[14];
7. **Off-topic**: binary value, with 1 referring to off-topic online posts that are political but not related to the budget;
8. **Language**: numerical value, with 0 referring to online posts in English, 1 referring to posts in Maltese, 2 referring to Maltese-English (Maltenglish) code-switched[15] posts, and 3 referring to posts in other languages.

The following is an example of an online post and the annotations for each annotation type:

Online post	"Online Maltese language spellchecker to be commissioned #maltabudget20"
Annotation Types	**Subjectivity**: 0; **Sentiment Polarity**: Positive; **Emotion**: Joy; **Sarcasm**: 0; **Irony**: 0; **Negation**: 0; **Off-topic**: 0; **Language**: 0

3.3 Quality

To ensure that the final datasets provided are of good quality, some basic pre-processing was carried out on the source data collected, whereas inter-rater reliability was calculated to determine that the level of agreement between the raters' annotations for each annotation type.

Pre-processing Basic pre-processing was carried out on the data collected as discussed in Section 3.1. Any deleted comments (by the respective newswire or comment owner) or comments that consisted of images only were removed from the newswires data, whereas any duplicate tweets and retweets were removed from the Twitter data. Moreover, any HTML tags and line breaks were also removed from the collected tweets.

Inter-rater Reliability The quality of the three datasets for each annotation type (described in Section 3.2 above), is evaluated through inter-rater reliability, that is, the level of agreement between the raters' annotations. The percent agreement (% Agree) is primarily calculated on the annotations performed by the three raters, which basic measure is calculated for two different levels, annotations agreed by all of the three raters (% Agree - 3 raters) and annotations agreed by two raters (% Agree - 2 raters). Two de facto statistical measurements, Fleiss' kappa (Fleiss and Cohen 1973) and Krippendorff's Alpha (Krippendorff 2011) have also been calculated. Fleiss' kappa takes chance agreement into consideration, which is commonly used for categorical variables, whereas Krippendorff's Alpha is used for content analysis to identify the agreement between raters and can apply to incomplete or missing data, any number of raters, any number of measurement level (nominal, ordinal, interval, ratio, etc.), and small and large sample sizes alike. Therefore, both measures are applicable when three or more raters perform the annotations and are used to measure the degree of agreement in classification over agreement that is expected when raters randomly assign class labels i.e., by chance.

Tables 3, 4, and 5 provide the inter-rater reliability agreement scores of the 2018, 2019 and 2020 Malta Government Budgets respectively, for each annotation type.

Annotation Type	**% Agree - 3 raters**	**% Agree - 2 raters**	**Fleiss' kappa**	**Krippendorff's Alpha**
Subjectivity	0.9841	0.0159	0.9776	0.9776
Sentiment Polarity	0.8978	0.1022	0.8721	0.8721
Emotion	0.4599	0.5401	0.5160	0.5001
Sarcasm	0.9804	0.0196	0.7626	0.7625
Irony	0.9818	0.0182	0.8256	0.8256
Negation	0.9300	0.0700	0.7539	0.7537
Off-topic	0.9370	0.0630	0.8227	0.8226
Language	1	0	1	1

Table 3: Malta Government Budget 2018 - Inter-rater reliability measures for each annotation type

Annotation Type	**% Agree - 3 raters**	**% Agree - 2 raters**	**Fleiss' kappa**	**Krippendorff's Alpha**
Subjectivity	1	0	1	1
Sentiment Polarity	0.7323	0.2677	0.7155	0.7151
Emotion	0.3804	0.6196	0.4269	0.4155
Sarcasm	0.9996	0.0004	0.9950	0.9950
Irony	0.9417	0.0583	0.6397	0.6394
Negation	0.9275	0.0725	0.6361	0.6353
Off-topic	0.9154	0.0846	0.8263	0.8263
Language	0.9175	0.0825	0.8714	0.8714

Table 4: Malta Government Budget 2019 - Inter-rater reliability measures for each annotation type

Annotation Type	**% Agree - 3 raters**	**% Agree - 2 raters**	**Fleiss' kappa**	**Krippendorff's Alpha**
Subjectivity	1	0	1	1
Sentiment Polarity	0.7351	0.2649	0.7131	0.7128
Emotion	0.4795	0.5205	0.5212	0.5159
Sarcasm	0.9990	0.0010	0.9827	0.9827
Irony	0.9326	0.0674	0.6167	0.6150
Negation	0.9559	0.0441	0.8833	0.8833
Off-topic	0.9020	0.0980	0.7881	0.7880
Language	0.9984	0.0016	0.9969	0.9969

Table 5: Malta Government Budget 2020 - Inter-rater reliability measures for each annotation type

Interpretation of the reliability results listed in the tables above differs between measures. All result values range from 0 to 1, where 0 signifies a perfect disagreement and 1 a perfect agreement for all measures. The % Agree is straightforward and the results simply provide an overview

[14] A negated post refers to the opposite of what is conveyed due to certain grammatical operations, such as 'not' (English) and 'mhux' ('not' in Maltese)

[15] Code-switching is a linguistic phenomenon that occurs when two or more languages are used in a single sentence or discourse

of the annotations that were in agreement by all three and two raters respectively. On the other hand, Fleiss' kappa results generally can be interpreted according to the classification guidelines by (Landis and Koch 1977) for categorical data. Such results are interpreted as follows: less than 0 - poor agreement, 0.0 to 0.20 - slight agreement, 0.21 to 0.40 - fair agreement, 0.41 to 0.60 - moderate agreement, 0.61 to 0.80 - substantial agreement, and 0.81 to 1.0 - almost perfect agreement. Lastly, Krippendorff's Alpha results are generally interpreted as follows: a value of 0.80 or higher constitutes a marker of good reliability, whereas results within the 0.667 to 0.80 range allow for tentative conclusions to be drawn (Krippendorff 2018). It is worth noting that these guidelines are more strict than the ones drawn up by Landis and Koch (Landis and Koch 1977). Therefore, one has to interpret these results in accordance to the particular hypothesis that is being tested and the validity requirements established on the research results.

An almost perfect agreement was achieved across the three datasets for *subjectivity* and *language* annotations, whereas a substantial/almost perfect agreement was achieved for *sentiment polarity*, *sarcasm*, and *off-topic* annotations. The *emotion* annotation was consistent across, with a moderate agreement. Lastly, the *irony* and *negation* annotations produced substantial to almost perfect agreements.

The moderate and contrasting results across datasets highlight the challenge behind these annotations tasks, especially when determining the *emotion*, *irony* and *negation*. In fact, the % Agree - 2 raters of online posts from newswires is higher than that of online posts from social networking services (in this case Twitter), due to user-generated content in newswires being lengthier and hence more difficult to annotate. This is the opposite in the case of % Agree - 3 raters of online posts from social networking services, which agreement is higher than its equivalent for newswires. Emotions are very subjective and can differ from one person to another, therefore can be annotated in an inconsistent manner (Mohammad and Turney 2013). Also, people tend to confuse sarcasm for irony and vice-versa, and sometimes find their interpretation difficult (Van Hee 2017). Irony proved to be more challenging to annotate than sarcasm, probably due to irony being a more sophisticated form of communication and the different types of irony categories, such as verbal and situational (Reyes, Rosso, and Veale 2013).

Consolidation A computational linguist (fourth rater mentioned in Section 3.2), consolidated the annotations to create the three final datasets. In cases where at a minimum two out of three raters agreed on the annotation, this was selected as being final. However, in cases of non-agreement between the three raters, the computational linguist discussed the results with the three raters and selected the most appropriate annotation value after an agreement was reached. This was only necessary for annotations containing categorical values, namely, *sentiment polarity*, *emotion*, and *language*.

4 Dataset Statistics and Discussion

The three datasets consist of 6,387 online posts in total. The distribution of the 2018, 2019, and 2020 datasets' annotations for the information (annotation types) discussed in Section 3.2 are presented in the following sub-sections. Moreover, any observations made during the annotation, quality, consolidation, and analysis processes are also discussed.

4.1 Subjectivity

Descriptive Statistics Table 6 presents the distribution of subjectivity annotations of the online posts for the 2018, 2019, and 2020 budget datasets. The majority of the posts in 2019 and 2020 datasets are subjective, with the ones in 2018 being mostly objective.

Subjectivity	**Budget 2018**	**Budget 2019**	**Budget 2020**
Subjective	38.66%	58.59%	58.32%
Objective	61.34%	41.41%	41.68%

Table 6: Distribution of subjectivity annotations

Observations

- Objective online posts can imply a sentiment polarity and emotion, since they can represent desirable or undesirable facts in certain specific domains or contexts (Liu 2015), such as the socio-economic domain in the context of the Government Budget. This opposes a general misconception within the OM research area, that objective text does not have any sentiment polarity by definition.
- Online posts that had a budget measure written (objective) followed by a subjective hashtag (see Example 1) were classified as being *objective*, since the emphasis was on the budget measure and not the subjective hashtag #WeNowLook2TheFuture.

  In first legislature Government managed to reduce debt from 70% of GDP, to 57.6%. #MaltaBudget18 #WeNowLook2theFUTURE **(Example 1)**

- Online posts that had a reference to the budget through a subjective hashtag, such as #WeNowLook2TheFuture (see Example 2), or a personal opinion followed by a subjective hashtag (see Example 3), were classified as *subjective*, since both instances emphasised the user's opinion.

  #MaltaBudget18 #WeNowLook2TheFUTURE #Malta **(Example 2)**

  A budget with a true socialist heart - grazzi @JosephMuscat_JM #MaltaBudget18 #WeNowLook2TheFuture **(Example 3)**

- The majority of online posts from Twitter (tweets) are *objective*, with their text referencing the budget measures being read out by the Minister for Finance.
- Even though retweets of objective budget measure online posts can indicate a a show of support/statement of approval, hence opinion of the entity retweeting, they were still classified as being *objective*, since they were posted

for information sharing purposes rather than for expressing their opinion through additional text (see Example 4).

RT @MaltaGov: Robotic surgery to be introduced in oncology #maltabudget20 **(Example 4)**

4.2 Sentiment Polarity

Descriptive Statistics Table 7 presents the distribution of sentiment polarity annotations of the online posts for the 2018, 2019, and 2020 budget datasets. All three datasets provide a high number of positive posts, with the negative ones showing an increase in each subsequent budget.

Sentiment Polarity	Budget 2018	Budget 2019	Budget 2020
Positive	63.21%	49.96%	53.86%
Neutral	20.82%	30.87%	23.02%
Negative	15.97%	19.17%	23.12%

Table 7: Distribution of sentiment polarity annotations

Observations

- Certain online posts contained multiple sentiment polarities, such as a *positive* polarity on the current budget or budget measure and a *negative* polarity on the future anticipated long-term effect of the said budget or budget measure. In these cases, the sentiment polarity is annotated in relation to the current budget or budget measure. Such instances highlight the importance of aspect-based OM and the challenges faced when determining an overall sentiment polarity to certain online posts.
- The sentiment polarity annotated does not always reflect the sentiment towards a particular budget or budget measure, due to it being compared to previous budgets or budget measures. For example, certain online posts have a *negative* sentiment due to the current budget being compared with the 2011 budget which had introduced several new taxes and did not achieve certain targets, such as Gross Domestic Product (GDP) growth.
- Similarly, certain polarities were aimed at a particular person or person's reaction and not at the budget itself. Example 5 is of a *negative* sentiment due to a lack of recognition for the budget by the opposition leader.

  Inkredibbli! @adriandeliapn jibqa' jsostni li dan hu Gvern bla pjan... ma tgħallem xejn. [Maltese] / Incredible! @adriandeliapn keeps insisting that this Government has no plan... he didn't learn anything. [English] #maltabudget19 @JosephMuscat_JM **(Example 5)**
- Certain online posts from newswires have a *negative* sentiment aimed at the writer or opinion of the previous online post and not the budget itself, in which cases the sentiment towards the budget would be the opposite, that is, *positive* (see Example 6).

  L-Unions kollha jghejdu li hu tajjeb u int u xi erba ohra tghejdu li kien hazin . Min ihobb jeqred, jeqred jibqa .[Maltese] / All the Unions say that it is good and you and a few others say that it was bad . Who enjoys grumbling, keeps grumbling . [English] **(Example 6)**
- Even though some online posts are classified as having a *negative* sentiment, this does not mean that the Government is being directly criticised but merely, the overall impression of the budget measure in question is not good and certain proposals and suggestions are being made in the subsequent online posts. For example, two online posts from the 2020 budget discuss the *negative* environmental impact of a budget measure that offers a grant to cover part of the cost of buying a battery storage system for owners of photovoltaic panels. These two online posts are actually providing alternative solutions to this measure, such as offering different feed-in tariffs, that can leave a *positive* environmental and economic impact.
- Online posts of a sarcastic or ironic nature result in changing and/or influencing the overall sentiment conveyed.

4.3 Emotion

Descriptive Statistics Table 8 presents the distribution of emotion annotations of the online posts for the 2018, 2019, and 2020 budget datasets. All three datasets produce a high number of posts conveying joy and anticipation emotions.

Emotion	Budget 2018	Budget 2019	Budget 2020
Joy	43.42%	34.46%	47.07%
Trust	8.59%	5.44%	2.75%
Fear	0.61%	1.30%	1.50%
Surprise	1.96%	3.58%	2.70%
Sadness	2.15%	4.23%	5.24%
Disgust	8.31%	6.48%	8.81%
Anger	6.12%	4.88%	6.48%
Anticipation	28.85%	39.64%	25.45%

Table 8: Distribution of emotion annotations

Observations

- The annotation of the emotion *joy* for certain online posts does not always reflect the typical joyous nature as usually expressed by a person through the use of certain special characters (e.g., exclamation mark) or emoticons (e.g., smiley face). However, in context of the Government Budget domain, this emotion category is the closest towards annotating one of a positive nature, such as the announcement of a positive budget measure (see Example 7). On the other hand, the *trust* emotion was used for posts expressing support to the Government.

  #MaltaBudget18 Live — Taskforce set up to focus on implementation of #Blockchain National Strategy in #Malta **(Example 7)**
- In the context of objective online posts (mostly tweets), emotions are somewhat different than those for subjective ones, which generally reflect the person's emotions. For example, the *joy* emotion shall bring a better quality of life and/or concrete support from budget measures.
- The *anticipation* emotion was used a lot in objective online posts that provided either links publishing budget updates (see Example 8), or updates on the current budget and/or budget measures announced (see Example 9).

#maltabudget2019 #taxes #maltaindependent https://t.cozdjCZxleYw **(Example 8)**

More investment expected in Artificial Intelligence and Internet of Things #MaltaBudget20 **(Example 9)**

- Throughout the 2018, 2019, and 2020 budget datasets, there were a number of online posts where the three annotators chose different emotion classifications, due to numerous emotions being expressed (e.g., *fear*, *anger*, and *sadness*) or the emotions being of a similar nature within the spectrum (e.g., *fear* and *surprise*).
- In certain instances the annotators found it challenging to select one of Plutchik's eight primary emotions, which highlights the complexity of such a task and the identification of the appropriate emotion category, which can easily differ from one person to another. Our claim is supported by (Susanto et al. 2020) where the authors highlight that "emotions are still a rather mysterious subject to study". This is reflected by a lack of universal emotion categorisation model and hence why numerous emotion classifications have been published in literature over the years.

4.4 Sarcasm

Descriptive Statistics Table 9 presents the distribution of sarcasm annotations of the online posts for the 2018, 2019, and 2020 budget datasets. The number of sarcastic posts is more or less consistent for the three datasets, which number diminishes from the 2018 till the 2020 budget.

Sarcasm	**Budget 2018**	**Budget 2019**	**Budget 2020**
Sarcastic	3.17%	2.98%	2.07%
Not Sarcastic	96.78%	97.02%	97.93%

Table 9: Distribution of sarcasm annotations

Observations

- Certain sarcastic online posts still keep their original sentiment polarity, e.g., *negative* (see Example 10).

Said like the true monkey that you are **(Example 10)**

- Other sarcastic online posts have a particular sentiment polarity even though in reality they convey an opposite one. Example 11 has a *positive* sentiment polarity due to its sarcastic nature and use of "face with tears of joy" emoticons (also present in original text), however, it coveys a *negative* one towards the referenced person.

Thanks for the advice pycho Joe!!! **(Example 11)**

4.5 Irony

Descriptive Statistics Table 10 presents the distribution of irony annotations of the online posts for the 2018, 2019, and 2020 budget datasets. The number of ironic posts slightly increased in the 2019 and 2020, when compared to 2018.

Observations

- Certain terms that are usually used to express a *positive* sentiment polarity, such as, "thanks", "kind", "hope" and "entertaining", have sometimes been used in ironic online posts, even though they are conveying the opposite meaning of what is being said towards the particular entity, such as a person (see Example 12).

thanks for the kind words Albie. **(Example 12)**

- The ironic nature of certain posts convey a *negative* sentiment polarity, however express a positive emotion, such as *joy*. Example 13 is in reality praising the existing Government for the positive budget measures being announced (e.g., incentives) and referring to previous Governments in a negative sentiment based on past budget measures (e.g., additional taxes).

How boring, he keeps using the same words - give, giving, we give. We are so accustomed, for 27 years, to hearing the words - pay, taxes, tariffs. sacrifices etc **(Example 13)**

Irony	**Budget 2018**	**Budget 2019**	**Budget 2020**
Ironic	3.78%	5.87%	5.81%
Not Ironic	96.22%	94.13%	94.19%

Table 10: Distribution of irony annotations

4.6 Negation

Descriptive Statistics Table 11 presents the distribution of negation annotations of the online posts for the 2018, 2019, and 2020 budget datasets. The 2020 budget dataset produced the highest number of negations.

Negation	**Budget 2018**	**Budget 2019**	**Budget 2020**
Negated	12.65%	8.72%	14.05%
Not Negated	87.35%	91.28%	85.95%

Table 11: Distribution of negation annotations

Observations

- Any negations within an online post that were not related to the budget or a particular budget measure announced by the Government for the respective year, were ignored. Example 14 contains a word ("never") that indicates a negation, however, it was ignored (in line with the annotation guidelines), since only the first sentence was related to the 2019 budget.

Yeah sure, a rise of 2 euros a week just to cover the extra cost of bread and milk, As for the other extra costs, we will tackle that in our next budget. Accepting more than 40,000 economic immigrants can never improve the way of life of the Maltese workers, JMO. **(Example 14)**

4.7 Off-topic

Descriptive Statistics Table 12 presents the distribution of off-topic annotations of the online posts for the 2018, 2019, and 2020 budget datasets. A substantial percentage of posts in each dataset are off-topic, especially the one from 2019.

Off-topic	Budget 2018	Budget 2019	Budget 2020
Off-topic	12.65%	19.73%	17.73%
On-topic	87.35%	80.27%	82.27%

Table 12: Distribution of off-topic annotations

Observations

- Online posts from newswires have a tendency to end up being classified as being off-topic due to several reasons, such as reference to previous Government administrations, measures e.g., pensions schemes introduced in the past, context of a previous online post misunderstood or its reply being ironic and not within context of Maltese politics, direct reference to a political figure after providing budget feedback e.g., opposition leader at the time, comparison of non-budget matters with other jurisdictions e.g., United Kingdom, and reference to current situations in the country e.g., scandals.
- Certain on-topic online posts (correctly annotated) indirectly refer to budget topics/measures, such as cost of bread and milk (see Example 15) which are related to the cost of the living allowance (COLA)[16], even though this is not specifically mentioned within the text.

Milk and bread in Malta cost nothing. **(Example 15)**

4.8 Language

Descriptive Statistics Table 13 presents the distribution of language annotations of the online posts for the 2018, 2019, and 2020 budget datasets. The majority of the posts in each dataset are in English, with the Maltese language (either as the primary language or as a secondary language) used in around a quarter of the posts.

Language	Budget 2018	Budget 2019	Budget 2020
English	71.52%	71.55%	79.99%
Maltese	4.34%	6.22%	3.21%
Maltese-English	23.20%	21.24%	15.97%
Other	0.93%	0.99%	0.83%

Table 13: Distribution of language annotations

- A total of 20 online posts from the 2018 budget dataset were classified as being written in other languages, namely: 15 in English-Italian, 1 in Italian, 1 in Maltese-Italian, 1 in English-French, 1 in Spanish-English, and 1 in Maltese-English-Italian.
- A total of 23 online posts from the 2019 budget dataset were classified as being written in other languages, namely: 12 in English-Italian, 3 consisted of links only, 2 in Dutch, 2 in English-French, 1 in Maltese-Italian, 1 in English-Spanish, 1 in English-Swedish, and 1 consisted only of one emoticon.
- A total of 16 online posts from the 2020 budget dataset were classified as being written in other languages, namely: 5 in English-Italian, 4 in Maltese-Italian, 1 in English-Spanish, 1 in English-French, 1 in English-Japanese, 1 in Maltese-English-Italian, 1 in Italian, 1 in Japanese, and 1 consisted only of emoticons.

Observations

- Most of the Maltese-English code-switched online posts result in the majority of the terms being in Maltese, with only a few words written in English. Some common occurrences are: "budget" (English) instead of "baġit" (Maltese), and "euro" (English) instead of "ewro" (Maltese). However, there were still some cases where it was the opposite, that is, English being the primary language and Maltese the secondary language.
- Several online posts were written in Maltese, however, they used a hashtag (tweets) written in English (e.g., #maltabudget20), hence were classified as being code-switched. This was a common occurrence across the three datasets obtained from Twitter.
- Loan words such as "cappuccino" are not of a Maltese origin, however they have been incorporated within our language (and others worldwide, such as English), therefore, they were not classified as being non-Maltese.
- Certain terms are well accepted in Maltese e.g., "amen" (Christian word spelling), however, there is a Maltese translation of this word "ammen"[17] and more so a Maltese version "hekk ikun"[18]. Therefore, given that this term is used worldwide, in principle classification in Maltese or English are both considered as being correct.
- The words "pastizz" (singular) or "pastizzi" (plural) refers to a traditional Maltese savoury pastry. Both words are in Maltese and have an English translation depending on the flavour (cheese cakes/pea cakes). Given that the words are widely used by the general public irrespective of the language, there were some instances where the annotators did not always recognise the words as being in Maltese and therefore classified the language as English, in cases where all the other text was in English.
- Acronyms/slang words referring to English phrases, such as LOL (laughing out loud), were treated in their original language during annotation. Therefore, online posts in Maltese containing such terms, were classified as Maltese-English.
- Hashtags such as "#MaltaSuccess", could have been meant to be in Maltese. However, due no Maltese characters used, the word "suċċess" (Maltese) might have been written as "success" (English). In this case, the words were classified as being in English.
- The level of inter-rater reliability agreement between Maltese and Maltese-English might be a bit lower than expected. This is due to the fact that certain terms, such as "man" within Maltese text and "pastizzi" within English text, have been embedded in our language and day-to-day vocabulary for such a long time, that they may have seemed natural for the annotators.

[16]https://www.gemma.gov.mt/cost-of-living-increase/

[17]https://www.independent.com.mt/articles/2015-09-01/blogs-opinions/Chiselling-the-Maltese-Language-6736141379

[18]https://www.timesofmalta.com/articles/view/Amen-written-in-Maltese.377498

4.9 Data Sources

Descriptive Statistics Table 14 presents the distribution of all annotations from the online posts for the consolidated 2018, 2019, and 2020 budget datasets, for each data source, namely Times of Malta (TOM), MaltaToday (MT), The Malta Independent (TMI) and Twitter (TW).

Data source	TOM	MT	TMI	TW
Total	869	761	93	4664
Subjectivity				
Subjective	864	759	93	875
Objective	5	2	0	3789
Sentiment Polarity				
Positive	164	141	10	3235
Neutral	188	114	17	1286
Negative	517	506	66	143
Emotion				
Joy	62	93	5	2476
Trust	74	42	4	243
Fear	22	29	5	16
Surprise	84	53	8	32
Sadness	110	65	10	60
Disgust	204	228	24	42
Anger	166	148	28	27
Anticipation	147	103	9	1768
Sarcasm				
Sarcastic	59	106	7	6
Not Sarcastic	810	655	86	4658
Irony				
Ironic	160	125	12	32
Not Ironic	709	636	81	4632
Negation				
Negated	273	211	33	227
Not Negated	596	550	60	4437
Off-topic				
Off-topic	429	504	24	113
On-topic	440	257	69	4551
Language				
English	586	470	58	3618
Maltese	117	134	9	39
Maltese-English	156	145	25	971
Other	10	12	1	36

Table 14: Distribution of annotations by data source

Observations Online posts from newswires tend to be of a more *negative* sentiment polarity to those from social networking services, such as Twitter, which are mostly of a *positive* sentiment polarity. The same applies for emotions, with the ones of a *positive* nature, such as *joy* and *trust*, being mostly present in Twitter, as opposed to the ones of a *negative* nature, such as *fear*, *sadness*, *disgust*, and *anger*, conveyed in online posts from newswires. A high number of online posts from Twitter (tweets) are *objective* due them being about budget measures (factual). In terms of sarcasm and/or irony, online posts of this nature are mostly found in newswires. Moreover, a large portion of tweets were carried out by the members of the Cabinet of Malta, therefore may not provide a true reflection of the general population. A similar observation was made by Mellon and Prosser in their political science study (Mellon and Prosser 2017). However, these online posts are still relevant since Twitter is an open social media platform that can be used by the general public, which social media data provides several opportunities for studying public opinion (Mellon and Prosser 2017).

4.10 Online Posts

Descriptive Statistics Tables 15, 16, 17 and 18 present statistics on the online posts for each data source, in terms of maximum, minimum, and average characters, and words within posts for each of the three datasets. Moreover, an analysis was carried out on emoticons/emojis in terms of total online posts containing at least one, overall total number for each dataset, and the highest and lowest number of emoticons/emojis present in a post for each dataset.

Data source	Budget 2018	Budget 2019	Budget 2020
Characters - Average	169.79	185.94	220.51
Characters - Maximum	1176	1851	1576
Characters - Minimum	5	4	10
Words - Average	29.38	31.67	38.60
Words - Maximum	173	324	280
Words - Minimum	1	1	2
Emoticons/Emojis - Total posts	4	4	0
Emoticons/Emojis - Overall total	5	5	0
Emoticons/Emojis - Highest number	2	2	0
Emoticons/Emojis - Lowest number	1	1	0

Table 15: Online posts statistics - Times of Malta

Data source	Budget 2018	Budget 2019	Budget 2020
Characters - Average	139.33	198.75	181.98
Characters - Maximum	748	2204	1350
Characters - Minimum	8	2	4
Words - Average	23.89	34.29	31.60
Words - Maximum	121	398	238
Words - Minimum	1	1	1
Emoticons/Emojis - Total posts	6	30	8
Emoticons/Emojis - Overall total	9	145	46
Emoticons/Emojis - Highest number	3	9	30
Emoticons/Emojis - Lowest number	1	1	1

Table 16: Online posts statistics - MaltaToday

Observations Online posts within newswires data sources tend to be much longer than ones made on social networking services, such as Twitter. This is evident from the statistics presented, with the largest post from the three datasets containing 2204 characters / 398 words (Budget 2019 - MaltaToday), whereas the largest post from Twitter consisted of 352 characters / 49 words (Budget 2019). Please note that

Data source	Budget 2018	Budget 2019	Budget 2020
Characters - Average	339.39	275.60	181.03
Characters - Maximum	2037	610	694
Characters - Minimum	15	13	31
Words - Average	57.39	50	30.90
Words - Maximum	368	124	108
Words - Minimum	2	2	5
Emoticons/Emojis - Total posts	1	0	0
Emoticons/Emojis - Overall total	1	0	0
Emoticons/Emojis - Highest number	1	0	0
Emoticons/Emojis - Lowest number	1	0	0

Table 17: Online posts statistics - The Malta Independent

Data source	Budget 2018	Budget 2019	Budget 2020
Characters - Average	108.87	117.56	134.23
Characters - Maximum	295	352	318
Characters - Minimum	23	14	34
Words - Average	13.91	14.62	17.48
Words - Maximum	49	49	49
Words - Minimum	1	1	1
Emoticons/Emojis - Total posts	55	31	100
Emoticons/Emojis - Overall total	88	60	296
Emoticons/Emojis - Highest number	5	10	14
Emoticons/Emojis - Lowest number	1	1	1

Table 18: Online posts statistics - Twitter

the maximum number of text content of a tweet can contain up to 280 characters (updated from 140 characters in November 2017). However, content returned by the TwitterAPI Python library includes certain links (e.g., of images embedded in a tweet) and certain character entity references (e.g., *&* for the & character), which content is made available for the end-users to decide on whether further pre-processing is needed, depending on their application. Moreover, the use of emoticons/emojis in online posts within social networking services (such as Twitter), is usually higher than those made in newswires. However, an online post from a newswire (Budget 2020 - MaltaToday) contained the highest number of emoticons/emojis (30).

4.11 Overall

- **Implicit vs. explicit opinions**: Certain online posts express a particular sentiment polarity and emotion, in view of the opinion expressed by someone else. Hereby, the users making such a post implicitly approve of the budget even though their post does not explicitly express it. Example 16 conveys a *negative* sentiment polarity and *sadness* emotion for the view expressed by Adrian Delia and not the respective budget.

> Opposition leader @adriandeliapn is as inept as they come. His analysis of #maltabudget18 is totally out of sync with people. **(Example 16)**

- **Sarcasm vs. Irony:** It is important to clarify that an online post can only be annotated as being either sarcastic or ironic and not both.
- **Aspect-based OM**: In certain cases, online posts are long in nature and contain opinions on multiple budget measures and/or Government entities.

> @adriandeliapn said that #maltabudget20 has no measures for women. Less tax for everyone, climate change measures, free transport for youths and elderly, minimum wage for severely disabled, higher pensions. This budget is for men and women. Stop putting people in isolated boxes **(Example 17)**

In Example 17, multiple budget measures (highlighted) and other aspects were mentioned. These have a different sentiment polarity (*positive* for each measure and *negative* for Adrian Delia) and emotion (*joy* for each measure and *anger* for Adrian Delia). This shows why aspect-based OM is important and the benefits of having such an approach that can interpret opinions in an accurate manner based on each aspect and/or entity.

- The **Government Budget** context within the **socio-economic** domain, is a complex topic of choice and this can be seen from the classification of certain on-topic/off-topic online posts which are not always straightforward to determine, especially given that user-generated content can be within the Government context but not within the context of the specific budget.

5 FAIR

The datasets adhere to the FAIR principles as follows:

- **Findable**: publicly available through the Zenodo[19] open-access repository;
- **Accessible**: through the Digital Object Identifier (DOI)[20] assigned by Zenodo;
- **Interoperable**: data available in a structured, open and machine-readable format, as comma separated values (CSV) files;
- **Reusable**: published under the Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International (CC BY-NC-SA 4.0) license[21] for non-commercial use.

The three datasets do not contain any sensitive data, since they only include published public user-generated content. The identity of the users has been protected, where no usernames have been provided with respect to online posts collected from newswires. As for Twitter, the Developer Agreement and Policy[22] shall be observed for all the data gathered. Therefore, only the Twitter IDs and respective annotation

[19]https://www.zenodo.org/

[20]https://www.doi.org/10.5281/zenodo.4650232

[21]https://www.creativecommons.org/licenses/by-nc-sa/4.0/

[22]https://developer.twitter.com/en/developer-terms/agreement-and-policy

types shall be distributed, which data can only be used for non-commercial research purposes.

6 Conclusion and Potential Applications

The three datasets provide a valuable resource for developing OM tools that gather political and socio-economic insights from user-generated content in Malta's two official languages, Maltese and English. These can be used by the Government of Malta for policy formulation, policy-making, decision-making, and decision-taking. Moreover, their use can support similar initiatives in other countries (e.g., Irish Government Budget), studies in the socio-economic domain and other application areas, such as Politics, Finance, Marketing, Advertising, Sales and Education.

Furthermore, these quality datasets are valuable for multiple research applications, namely:

- Tools and resources for low-resourced languages, such as Maltese;
- NLP for social media content in Maltese and English;
- NLP approaches for the analysis and processing of mixed-language online user-generated content, with a focus on code-switching in Maltese.
- OM on monolingual (English/Maltese) and code-switched online user-generated content;
- Aspect-based OM for multiple social opinion dimensions;
- Fine-grained opinion search and opinion summarisation;
- Subjectivity detection, sentiment analysis, emotion analysis, sarcasm detection, and irony detection (as separate research areas or otherwise) in multiple application areas.

Acknowledgments

This work is funded by the ADAPT Centre for Digital Content Technology which is funded under the SFI Research Centres Programme (Grant 13/RC/2106).

References

Charalabidis, Y.; Maragoudakis, M.; and Loukis, E. 2015. Opinion mining and sentiment analysis in policy formulation initiatives: The EU-community approach. In *International Conference on Electronic Participation*, 147–160. Springer.

Cortis, K.; and Davis, B. 2019. A Social Opinion Gold Standard for the Malta Government Budget 2018. In *Proceedings of the 5th Workshop on Noisy User-generated Text (W-NUT 2019)*, 364–369.

Cortis, K.; and Davis, B. 2020. Over a Decade of Social Opinion Mining. *arXiv preprint arXiv:2012.03091* .

Fleiss, J. L.; and Cohen, J. 1973. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educational and psychological measurement* 33(3): 613–619.

Kemp, S. 2019. Digital 2019 Malta. URL https://www.slideshare.net/DataReportal/digital-2019-malta-january-2019-v01. DataReportal.

Krippendorff, K. 2011. Computing Krippendorff's alpha-reliability. https://repository.upenn.edu/asc_papers/43. Accessed: 2011-01-25.

Krippendorff, K. 2018. *Content analysis: An introduction to its methodology*. Sage publications.

Landis, J. R.; and Koch, G. G. 1977. The measurement of observer agreement for categorical data. *biometrics* 159–174.

Liu, B. 2010. Sentiment analysis and subjectivity. In *Handbook of Natural Language Processing, Second Edition. Taylor and Francis Group, Boca*.

Liu, B. 2015. *Sentiment Analysis: Mining Opinions, Sentiments, and Emotions*. Cambridge University Press. doi: 10.1017/CBO9781139084789.

Martin, I. 2020. 84% say Times of Malta is their main source of online news. https://timesofmalta.com/articles/view/times-of-malta-is-countrys-most-popular-news-site-and-paper-survey.800391. Times of Malta.

Mellon, J.; and Prosser, C. 2017. Twitter and Facebook are not representative of the general population: Political attitudes and demographics of British social media users. *Research & Politics* 4(3): 2053168017720008.

Mohammad, S. M.; and Turney, P. D. 2013. Crowdsourcing a word–emotion association lexicon. *Computational Intelligence* 29(3): 436–465.

Plutchik, R. 1980. Chapter 1 - A GENERAL PSYCHO-EVOLUTIONARY THEORY OF EMOTION. In Plutchik, R.; and Kellerman, H., eds., *Theories of Emotion*, 3 – 33. Academic Press. ISBN 978-0-12-558701-3. doi:https://doi.org/10.1016/B978-0-12-558701-3.50007-7.

Pustejovsky, J.; and Stubbs, A. 2012. *Natural Language Annotation for Machine Learning: A guide to corpus-building for applications*. " O'Reilly Media, Inc.".

Reyes, A.; Rosso, P.; and Veale, T. 2013. A multidimensional approach for detecting irony in twitter. *Language resources and evaluation* 47(1): 239–268.

Rosner, M.; Joachimsen, J.; Rehm, G.; and Uszkoreit, H. 2012. *The Maltese language in the digital age*. Springer.

Schembri, S. 2019. Malta: The Ultimate AI Launchpad - A Strategy and Vision for Artificial Intelligence in Malta 2030. https://malta.ai/wp-content/uploads/2019/10/Malta_The_Ultimate_AI_Launchpad_vFinal.pdf. Parliamentary Secretariat for Financial Services, Digital Economy and Innovation, Office of the Prime Minister.

Susanto, Y.; Livingstone, A. G.; Ng, B. C.; and Cambria, E. 2020. The hourglass model revisited. *IEEE Intelligent Systems* 35(5): 96–102.

Van Hee, C. 2017. *Can machines sense irony?: exploring automatic irony detection on social media*. Ph.D. thesis, Ghent University.

Xuereb, M. 2020. 'We consulted online media' - Justice Minister on rule-of-law reform bills. https://timesofmalta.com/articles/view/we-consulted-social-media-justice-minister-on-rule-of-law-reform-bills.825589. Times of Malta.