Article

Text mining analysis to evaluate stakeholders’ perception on the welfare of equines, small ruminants and turkeys

Firstname Lastname 1, Firstname Lastname 2 and Firstname Lastname 2,\*

1 Affiliation 1; e-mail@e-mail.com

2 Affiliation 2; e-mail@e-mail.com

**\*** Correspondence: e-mail@e-mail.com; Tel.: (optional; include country code; if there are multiple corresponding authors, add author initials) +xx-xxxx-xxx-xxxx (F.L.)

Received: date; Accepted: date; Published: date

**Simple Summary:** It is vitally important that scientists are able to describe their work simply and concisely to the public, especially in an open-access on-line journal. The simple summary consists of no more than 200 words in one paragraph and contains a clear statement of the problem addressed, the aims and objectives, pertinent results, conclusions from the study and how they will be valuable to society. This should be written for a lay audience, i.e., no technical terms without explanations. No references are cited and no abbreviations. Submissions without a simple summary will be returned directly. Example could be found at http://www.mdpi.com/2076-2615/6/6/40/htm.

**Abstract:** A single paragraph of about 200 words maximum. For research articles, abstracts should give a pertinent overview of the work. We strongly encourage authors to use the following style of structured abstracts, but without headings: (1) Background: Place the question addressed in a broad context and highlight the purpose of the study; (2) Methods: Describe briefly the main methods or treatments applied; (3) Results: Summarize the article's main findings; and (4) Conclusions: Indicate the main conclusions or interpretations. The abstract should be an objective representation of the article, it must not contain results which are not presented and substantiated in the main text and should not exaggerate the main conclusions.

**Keywords:** animal welfare; stakeholder perception; text mining; horse; donkey; goat; sheep; turkey

1. Introduction

* Importanza dell’opinione degli stakeholders
* Uso dei questionari per valutare opinione stakeholders
* Problematiche dell’uso dei questionari
* Text mining analysis
* Progetto awin (approccio bottom up)
* Obiettivi

Concern for animal welfare is significantly increasing in many parts of the world [1].

Stakeholders, for example companion animal owners, farmers, horse owners and veterinary surgeons can have direct experience of the welfare issues which exist and consultation of stakeholders has been used to identify welfare problems in companion animals (for example [11]; dogs), farm animals (for example [12]; dairy cows) and equine species (for example [13]; working equids). Collins et al [14, 15] utilised a Delphi approach, whereby experts were consulted in three systematic, iterative rounds, to facilitate the identification of the main welfare problems facing horses in Ireland[2].

Sentiment analysis or opinion mining is the computational study of people’s opinions, appraisals, attitudes, and emotions toward entities, individuals, issues, events, topics and their attributes. The task is technically challenging and practically very useful [3].

The welfare of animals significantly depends on how stakeholders perceive their needs and behave with them. This work aimed at investigating stakeholder perception of the welfare of sheep, goats, horses, donkeys and turkeys using a text mining approach.

Text mining, also know ad text data mining (Hearst, M.A. 1997) or knowledge discovery from textual databases (Feldman, R & Dagan I. 1995) refers to the process of extracting interesting and non-trivial patterns of knowledge from unstructured text documents (ref)

2. Materials and Methods

2.1 Web-survey

For each species (horses, donkeys, turkeys, sheep and goats), a survey was developed and consisted of 14 open-ended questions (max 150 characters) referring to different aspects of animal welfare. To take the survey, participants were required to be over the age of 18. With the purpose of reaching more participants, the survey was translated in five languages (English, French, Italian, Portuguese and Spanish). Questions were related to four main topics: 1) needs (what does an animal need to be fit, healthy and productive?); 2) behaviour (how might an animal behave/react in response to the following situations: noise, isolation, presence of known/unknown animal/person?) 3) emotions (how might an animal feel in response to the following situations: noise, isolation, presence of known/unknown animal/person?); and 4) welfare indicators (looking at your neighbor's animals, what signs would you identify, to evaluate: accommodation, feeding, health, manifestation of normal and abnormal behaviour?). The survey was published on a web platform (<http://www.questionari.unimi.it/awin/>) and it was freely accessible for 15 months. The web link was shared by email, social network and hosted in web-sites of several academic and international organizations (such as FAO, International Society for Equitation Science, Italian Equestrian Federation, Istituto Zooprofilattico Sperimentale dell'Abruzzo e del Molise "Giuseppe Caporale").

**Table xxx.** This is a table. Tables should be placed in the main text near to the first time they are cited.

|  |  |
| --- | --- |
| **Question** | **Type of question** |
| In your opinion, what do animal need to be good, healthy and productive? | Open text (max 150 characters) |
| Looking at your neighbor's horse, what signs would you identify, to evaluate | |
| The conditions of accommodation | Open text (max 150 characters) |
| Feeding conditions | Open text (max 150 characters) |
| Health conditions | Open text (max 150 characters) |
| The manifestation of normal behavior | Open text (max 150 characters) |
| The manifestation of abnormal behavior | Open text (max 150 characters) |

1 Tables may have a footer.

2.2 Text mining analysis

The term text mining refers to a "process of distillation of useful information from a text". (T. Kwartler 2017): It is a set of quantitative methods that use the words present in a text as "units" of analysis. It applies to different types of texts: books, tweets, mails, open-ended surveys, etc. .. any text can in fact be analyzed.

The text is first "tokenized" ie reduced to a sequence of simple terms deprived of those words that serve the sensible and comprehensible definition of a period (article, avverb, number, punteggiatura, ecc…) . This document called Corpus is then transformed into a Term Document Matrix (TDM), that is a matrix that shows for each single term how many times it appears in a single document. From this matrix all the types of textual analysis are obtained, including: word frequency, word association, , cluster analysis, topic model analysis, sentiment analysis and many more.

In this study the text we are working on is the result of a Survey in which open-ended answers to definite questions have been given. For each question every single answer is defined as a 'document’. A pre-process consisting of: English translattion for other leanguage anwers, removal all terms according to a preselet list of so called “stop-words”, cleaning by numbers, puntuaction, was undertaken to get the Corpus and then the TDM for each single question. On TDM was performed first a simple frequency words analysis, whose results are shown as barplot, and then terms association analysis eas undertaken. In text mining analysis terms association is like correlation. It is refers to the term pairngs ( whrn the term x appears, the other term y is associated with it) and it is not directly related to frequency of terms. Unlike statistical correlation , association it is measured in a range between 0 and 1.  Scores range from 0 to 1. A score of 1 means that two words always appear together in documents, while a score approaching 0 means the terms seldom appear in the same document.As often as two terms appear in the same document, the stronger is their association. For each question the association study was performed between the most frequent word in the responses and all the other terms with a minimum frequency of occurrence of 10. Associations with a value greater than or equal to 0.20 were considered relevant.All analysis were performed in R, a freely available language and environment for statistical computing and graphics (R Development Core Team (2008), by “tm” package (Ingo Feinerer and Kurt Hornik (2018)). tm: Text Mining Package. R package version 0.7-6. [https://CRAN.R-project.org/package=tm](https://cran.r-project.org/package=tm)

Word frequency analysis

3. Results

A total of 270 surveys were properly filled out (122 for horses, 81 for sheep, 36 for goats, 18 for turkeys and only 13 for donkeys). The respondents came from 32 different Countries spread in the five continents, but the majority of them were from European countries (191), among which 41% were Italian stakeholders (Table XXX).

**Table xxx.** This is a table. Tables should be placed in the main text near to the first time they are cited.

|  |  |
| --- | --- |
| **Country** | **Number of respondents** |
| Europe | 191 |
| North America | 32 |
| South and Central America | 12 |
| Oceania | 25 |
| Asia | 7 |
| Africa | 4 |

1 Tables may have a footer.

In general, stakeholders were mainly veterinarians (113 participants), but some differences in stakeholders’ role were observed between species. For example, for horses most of the respondents were private owners and trainers (41 and 16, respectively), whereas for goats they were mainly farmers (19) and for turkeys contract farmers (2) were also included (Table…).

**Table xxx.** This is a table. Tables should be placed in the main text near to the first time they are cited.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Sheep (N=81)** | **Goats (N=36)** | **Horses (N=122)** | **Donkeys (N=13)** | **Turkeys (N=18)** |
| Veterinarian | 45 (56%) | 12 (33%) | 43 (35%) | 4 (23%) | 9 (50%) |
| Farmer | 19 (23%) | 17 (47%) | 8 (7%) | 8 (15%) | 5 (28%) |
| Contract farmer | 0 | 0 | 0 | 0 | 2 (11%) |
| Technician | 17 (21%) | 7 (19%) | 14 (11%) | 2 (62%) | 2 (11%) |
| Owner | 0 | 0 | 41 (34%) | 0 | 0 |
| Trainer | 0 | 0 | 16 (13%) | 0 | 0 |

1 Tables may have a footer.

Gender of participants was balanced in all species (Table X), except for horses (85% women). Most of the participants (77%) aged between 31 and 60 years.

**Table xxx.** This is a table. Tables should be placed in the main text near to the first time they are cited.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Sheep** | **Goats** | **Horses** | **Donkeys** | **Turkeys** |
| Male | 49 (60%) | 21 (58%) | 18 (15%) | 4 (31%) | 10 (56%) |
| Female | 32 (40%) | 15 (42%) | 105 (85%) | 9 (69%) | 8 (44%) |

1 Tables may have a footer.

3.1 In your opinion, what do animal need to be good, healthy and productive?

To describe welfare requirements, the words “food” and “water” were the most frequently used (68 and 40 times, respectively), meaning that participants considered the welfare principle Good Feeding as the most relevant for horses. As for the principle Good Housing, “shelter” and “exercise” appeared of primary importance (31 and 29 times, respectively), together with the presence of “pasture”, mentioned 18 times. The principle Good Health was linked to “care” (24) and “health” (20). To describe the principle Appropriate Behaviour, respondents used the words “proper training” (23) and “company”, mentioned only 16 times.



**Figure xx.** The graph report the frequency of words for the question “In your opinion, what do horse need to be good, healthy and productive?

The principle Good Housing appeared to be the most important for donkey stakeholders, the word “shelter” was mentioned 7 times. Then, we found the words “food” (6) and “water” (5) linked with the welfare principle Good Feeding. Interestingly, for donkeys the word “deworming” was mentioned twice.

3.2 Looking at your neighbor's horse, what signs would you identify, to evaluate accommodation, feeding, health, manifestation of normal and abnormal behaviour?

The results of the word frequency analysis for horses are reported in Figure 1. The most frequent word used by stakeholder to evaluate the conditions of accommodation (Figure 1 (a)) was “clean”, followed by “bedding”, “box”, “shelter” and “water”. The term “clean” was associated with the words: “floors” (0.33), “manger” (0.33), “drinkers” (0.32), “walls” (0.32) and “water” (0.31). An example of answer to the question “looking at your neighbor’s horse, what signs would you identify, to identify the condition of accommodation?” was: “Presence and cleanliness of bedding, repair and maintenance of buildings, pastures and equipment”.



**Figure 1.** The graphs report the frequency of words for each of the five questions (a) Looking at your neighbor’s horse, what signs would you identify, to evaluate the conditions of accommodation? (b). Looking at your neighbor’s horse, what signs would you identify, to evaluate the feeding conditions? (c) Looking at your neighbor’s horse, what signs would you identify, to evaluate health condition? (d) Looking at your neighbor’s horse, what signs would you identify, to evaluate the manifestation of normal behavior? (e) Looking at your neighbor’s horse, what signs would you identify, to evaluate the manifestation of abnormal behavior?

The words “roughage”, “water”, “quality” and “clean” were used by stakeholder to evaluate the feeding conditions (Figure 1 (b)). The term “roughage” was associated with the words: “pasture” (0.43) and “access” (0.39). An example of answer to the question “Looking at your neighbor’s horse, what signs would you identify, to evaluate the feeding conditions?” was: “A lot of roughage. Free access to clean hay that is not dusty or moldy. Or access to grazing”.

The health condition (Figure 1 (c)) was described with words such as “coat”, “body score”, “eyes” and “hooves”. The term “coat” was associated with the words: “cleaned” (0.26) and “worming” (0.26). An example of answer to the question “Looking at your neighbor’s horse, what signs would you identify, to evaluate health condition?“ was: “Body score and coat shine/condition. Horses posture, movement (lameness)”.

The most frequent word used by stakeholder to evaluate the manifestation of normal behaviour (Figure 1 (d)) was “group”, followed by “grazing”, “calm”, “relaxed” and “playing”. The term “group” was associated with the words: “happy” (0.38) and “curious” (0.30). An example of answer to the question “Looking at your neighbor’s horse, what signs would you identify, to evaluate the manifestation of normal behavior?” was: “horses are relaxed but still interested in their surroundings, they are eating hay or grazing, socializing together and allogrooming too”.

Finally, abnormal behavior (Figure 1 (e)) was described using words such as “weaving”, “aggression” and “stereotypies”. The term “weaving” was associated with the word “eat” (0.31). An example of answer to the question “Looking at your neighbor’s horse, what signs would you identify, to evaluate the manifestation of abnormal behavior?” was: “biting on doors, walls, posts. Kicking at walls. Constant or frequent vocalizing. Weaving, windsucking, pacing up and down, holding head high, eyes wide open, withdrawn, uninterested in surroundings, not interested in food, laying ears flat back when people approach, sudden or explosive movements”.

4. Discussion

Valutazione del campione e del numero dei partecipanti per specie

Coinvolte più figure professionali dello stesso settore

Authors should discuss the results and how they can be interpreted in perspective of previous studies and of the working hypotheses. The findings and their implications should be discussed in the broadest context possible. Future research directions may also be highlighted.

Horses (Q1): The results showed that horse stakeholders consider welfare primary linked with appropriate feeding. At the same time, allowing the possibility to interact with conspecifics and spend time at pasture is paramount to guarantee horse welfare. Stakeholders’ involvement is fundamental for any action intended to improve animal welfare; this work portrays the stakeholders’ perception, highlighting the need of proper dissemination of scientific knowledge.

5. Conclusions

This section is not mandatory, but can be added to the manuscript if the discussion is unusually long or complex.

**Supplementary Materials:** The following are available online at www.mdpi.com/xxx/s1, Figure S1: title, Table S1: title, Video S1: title.

**Author Contributions:** For research articles with several authors, a short paragraph specifying their individual contributions must be provided. The following statements should be used “conceptualization, X.X. and Y.Y.; methodology, X.X.; software, X.X.; validation, X.X., Y.Y. and Z.Z.; formal analysis, X.X.; investigation, X.X.; resources, X.X.; data curation, X.X.; writing—original draft preparation, X.X.; writing—review and editing, X.X.; visualization, X.X.; supervision, X.X.; project administration, X.X.; funding acquisition, Y.Y.”, please turn to the [CRediT taxonomy](http://img.mdpi.org/data/contributor-role-instruction.pdf) for the term explanation. Authorship must be limited to those who have contributed substantially to the work reported.

**Funding:** Please add: “This research received no external funding” or “This research was funded by NAME OF FUNDER, grant number XXX” and “The APC was funded by XXX”. Check carefully that the details given are accurate and use the standard spelling of funding agency names at https://search.crossref.org/funding, any errors may affect your future funding.

This survey was funded by the Animal Welfare Indicators (AWIN) Project, which has been co-financed by the European Commission, within the VII Framework Program (FP7-KBBE-2010-4, Grant n. 266213).

**Acknowledgments:** In this section you can acknowledge any support given which is not covered by the author contribution or funding sections. This may include administrative and technical support, or donations in kind (e.g., materials used for experiments).

CTU - the university eLearning and audiovisual production Center of Milan

Tutti quelli che hanno ospitato il link del questionario

**Conflicts of Interest:** “The authors declare no conflict of interest.”

Appendix A

The appendix is an optional section that can contain details and data supplemental to the main text. For example, explanations of experimental details that would disrupt the flow of the main text, but nonetheless remain crucial to understanding and reproducing the research shown; figures of replicates for experiments of which representative data is shown in the main text can be added here if brief, or as Supplementary data. Mathematical proofs of results not central to the paper can be added as an appendix.

Appendix B

All appendix sections must be cited in the main text. In the appendixes, Figures, Tables, etc. should be labeled starting with ‘A’, e.g., Figure A1, Figure A2, etc.

References

1. Buddle, E. A.; Bray, H. J.; Ankeny, R. A. “I Feel Sorry for Them”: Australian Meat Consumers’ Perceptions about Sheep and Beef Cattle Transportation. *Animals* **2018**, *8*, 1–13.

2. Horseman, S. V.; Buller, H.; Mullan, S.; Whay, H. R. Current Welfare Problems Facing Horses in Great Britain as Identified by Equine Stakeholders. *PLOS ONE* **2016**, *11*, e0160269.

3. Liu, B.; Zhang, L. A Survey of Opinion Mining and Sentiment Analysis. In *Mining Text Data*; Springer US: Boston, MA, 2012; pp. 415–463.

1. Author 1, A.B. Title of Thesis. Level of Thesis, Degree-Granting University, Location of University, Date of Completion.
2. Title of Site. Available online: URL (accessed on Day Month Year).

|  |  |
| --- | --- |
| copyRight | © 2019 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/). |

Ingo Feinerer and Kurt Hornik (2018)). tm: Text Mining Package. R package version 0.7-6. [https://CRAN.R-project.org/package=tm](https://cran.r-project.org/package=tm)

Ingo Feinerer, Kurt Hornik, and David Meyer (2008). Text Mining Infrastructure in R. Journal of Statistical Software 25(5): 1-54. URL: <http://www.jstatsoft.org/v25/i05/>.

R Development Core Team (2008). R: A language and environment for

statistical computing. R Foundation for Statistical Computing,

Vienna, Austria. ISBN 3-900051-07-0, URL [http://www.R-project.org.](http://www.r-project.org./)