IRISA

Institut de Recherche en Informatique et Systèmes Aléatoires

# Automatic Emotions Analysis for French Email Campaigns optimization

**Alexis Blandin** [1,2,4], Farida Saïd[1,3], Jeanne Villaneau[1,2], Pierre-François Marteau[1,2]

[1] *Université Bretagne Sud (France)*

[2] *IRISA*

[3] *LMBA*

[4] *UNEEK-Kosmopolead*

## Alexis BLANDIN

- PhD student since January 2020
  - Title : "*Opinion Analysis for Customer Relationship Optimization*"
  - Member of the IRISA laboratory : team **EXPRESSION**
  - CIFRE fellowship with **UNEEK**
- "*Age Recommendation for Texts*" - LREC 2020
- Graduate from **ENSSAT** engineering school : specialization in Artificial Intelligence
- Master degree in Computer Science from "*Université de Rennes 1*"

# Main goals

uneek Customer Relationship Management

How to help CRM users to manage their customer relationship through email?

- Analyze the emotions conveyed by the text of email campaigns
- Evaluate how these emotions affect the performance of newsletters

# Related work

## Related work – Email analysis

"Les chiens sont contents"

- Raw text

"Chien être content"

- Lemmatization and removing stop words

[0,0,0,0,0,0]
[0,0,0,0,0,0]
[0,0,0,1,0,0]

- Is the word providing an emotion?

[0,0,0,1/3,0,0]

- Calculate the sentence average

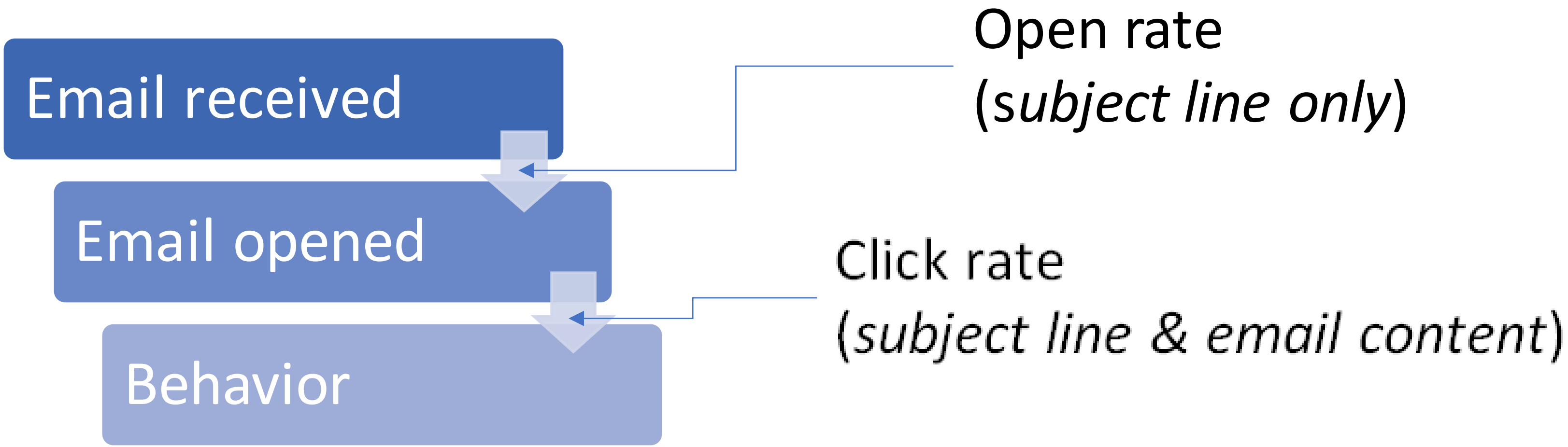
- A baseline approach that considers the 6 fundamental emotions: anger, fear, sadness, joy, disgust and surprise [P.Eckman,1999]
- We added 2 opinion scores: **polarity** and **subjectivity**

## Related work – Marketing studies

Emotion intended by the sender → Emotion conveyed by the email → Emotion perceived by the reader

- Lack of face-to-face communication and context
- **Neutrality** or **Negativity** bias on the perceived emotions [K.Byron,2008; V.Rodriguez et al., 2021]
- Does emotion have an impact on customer behavior?

## Related work - Email analysis




- Other measures of behavior: purchase rate, unsubscribe rate, etc. [R.Miller et al. , 2016]
- We focused on **open** and **click** rates: most relevant in our dataset

## Our work- Main Objectives

- **Analysis** of email campaigns using **emotion detection**

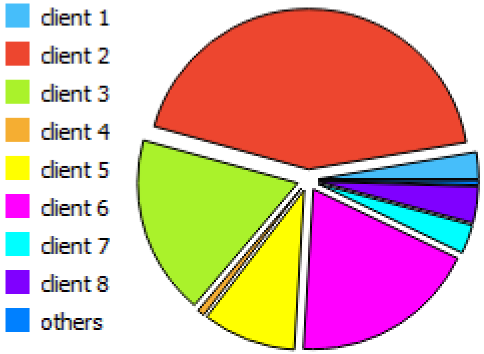
- Explore correlations **between newsletter** performance and **emotion embeddings**

- Test how these **correlations** can help **predict** newsletter **performance** based on **emotional tone**

- **More than 950 non-commercial** newsletters

- **We assume that** the emotions conveyed by the email do not depend on the sender

- Each newsletter is characterized by :
    - a subject line
    - a content text
    - an open rate
    - a click rate



# Pearson correlations

| Features | Open rate | Click rate |
|---|---|---|
| File size (FS) | **-0.14*** | **0.25*** |
| Subject line length (SL) | **-0.13*** | **0.18*** |
| Subject line polarity (SP) | -0.07** | -0.03$^{ns}$ |
| Subject line subjectivity (SS) | -0.01$^{ns}$ | -0.07* |
| Content Polarity (CP) | - | 0.09** |
| Content Subjectivity (CS) | - | -0.07* |
| Content Joy (J) | - | -0.10** |
| Content Fear (F) | - | **-0.11*** |
| Content Sadness (S) | - | **-0.23*** |
| Content Anger (A) | - | 0.06$^{ns}$ |
| Content Surprise (Su) | - | **-0.11*** |
| Content Disgust (D) | - | **-0.07*** |

*p-value $<$ .05, **p-value $<$ .01, ***p-value $<$ .001, $^{ns}$ not significant

Known in email features analysis

Little to not significant

Inverse associations

- How to determine **the optimal number of clusters** ?

| PCA[a] | Explained variance | Number of clusters[b] | silhouette score |
|---|---|---|---|
| 1 | 24% | 2 | 0.577 |
| 2 | 40% | 2 | 0.501 |
| 3 | 53% | 4 | 0.411 |
| 4 | 63% | 2 | 0.358 |
| 5 | 72% | 2 | 0.274 |
| 6 | 79% | 2 | 0.269 |
| 7 | 86% | 3 | 0.250 |
| 8 | 91% | 2 | 0.258 |
| 9 | 96% | 4 | 0.392 |
| 10 | 100% | 4 | 0.366 |

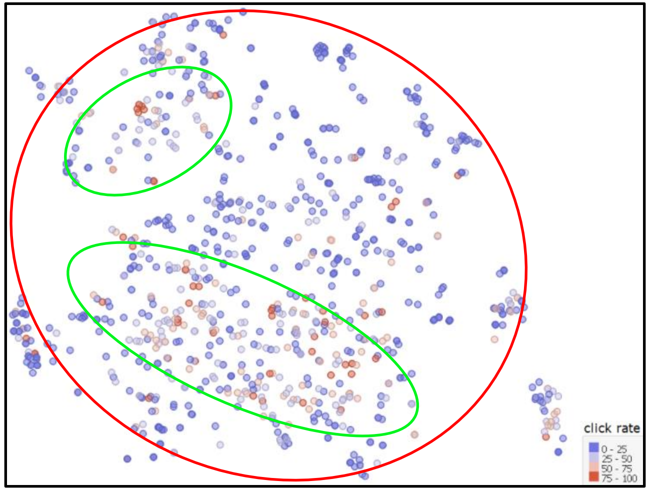
[a] Number of PCA components
[b] The optimal number of clusters is chosen to maximize the silhouette score

A good compromise

- Are there significant differences in the distribution of emotions between the **good (best half)** and **bad (worst half)** newsletters ?

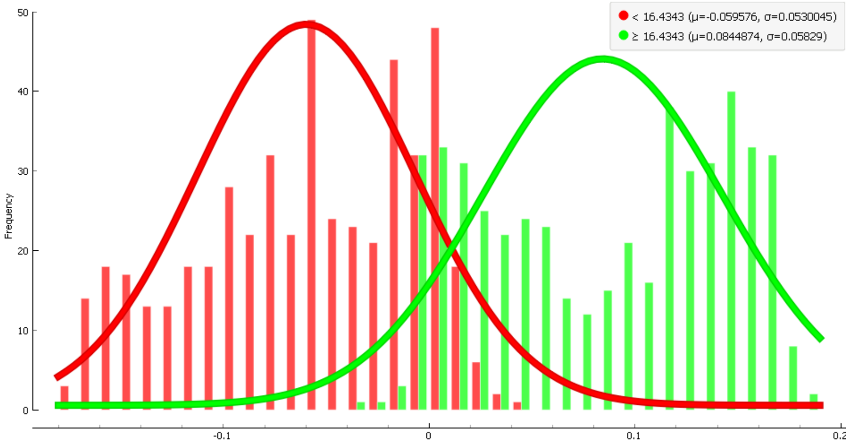
- **Bad** newsletters seem **evenly distributed**

- **Good** newsletters seem to **form clusters**

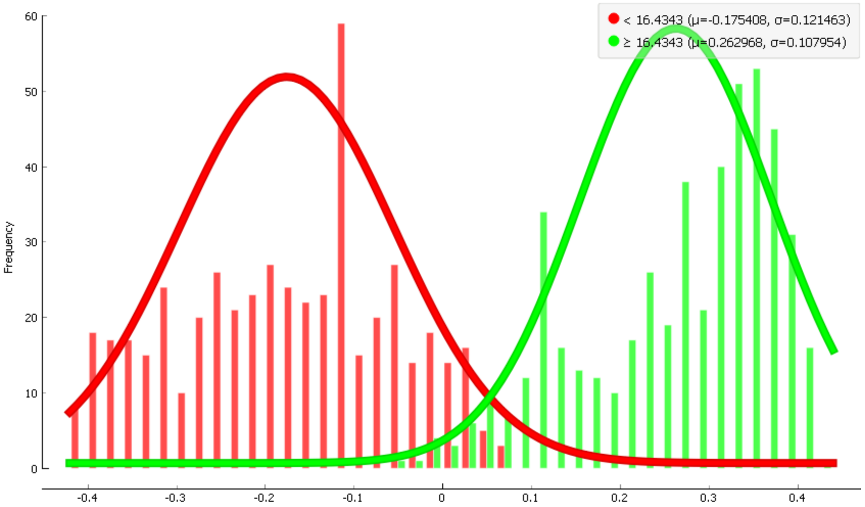



*t-SNE projection of our dataset*

# Our work – Silhouette Scores




*Silhouette score distribution with subject lines*




*Silhouette score distribution without subject lines*

# Our work – Supervised Classification and Features Selection

| Classifier | F1 Score | Precision | Recall |
|---|---|---|---|
| With subject line information | | | |
| **AdaBoost** | **0.723** | **0.724** | **0.724** |
| Neural Network | 0.712 | 0.712 | 0.712 |
| Random Forest | 0.711 | 0.711 | 0.711 |
| kNN | 0.681 | 0.688 | 0.683 |
| Naive Bayes | 0.666 | 0.666 | 0.666 |
| SVM | 0.607 | 0.617 | 0.612 |
| Logistic Regression | 0.585 | 0.594 | 0.590 |
| Constant | 0.500 | 0.500 | 0.500 |

| Without subject line information | | | |
|---|---|---|---|
| **Model** | **F1 Score** | **Precision** | **Recall** |
| **AdaBoost** | **0.722** | **0.723** | **0.723** |
| Neural Network | 0.714 | 0.715 | 0.715 |
| Random Forest | 0.710 | 0.710 | 0.710 |
| kNN | 0.679 | 0.683 | 0.680 |
| Naive Bayes | 0.666 | 0.666 | 0.666 |
| SVM | 0.628 | 0.640 | 0.633 |
| Logistic Regression | 0.621 | 0.643 | 0.630 |
| Constant | 0.500 | 0.500 | 0.500 |

| Feature | F1-score with a single feature | F1-score with all but one feature |
|---|---|---|
| Subject line polarity | 0.498 | 0.720 |
| Subject line subjectivity | 0.503 | 0.721 |
| Content Polarity | 0.614 | 0.719 |
| Content Subjectivity | 0.570 | 0.725 |
| Content Joy | 0.624 | 0.723 |
| Content Fear | 0.604 | 0.722 |
| **Content Sadness** | **0.633** | **0.711** |
| Content Anger | 0.618 | 0.715 |
| Content Surprise | 0.614 | 0.721 |
| Content Disgust | 0.626 | 0.721 |

## Conclusion and future work

- Emotions **influence** the performance of French email campaigns
- **Need for further study** to provide a writing recommender tool
- **Improve** our emotion detection analysis
- **Share our dataset** for reproducibility

# BIBLIOGRAPHY


[1] K. Byron, "Carrying too heavy a load? the communication and miscommunication of emotion by email, "The Academy of Management Review, vol. 33, no. 2, pp. 309–327, 2008. [Online]. Available: http://www.jstor.org/stable/20159399

[2] J.-E. Kim and K. Johnson, "The impact of moral emotions on cause-related marketing campaigns: A cross-cultural examination," Journal of Business Ethics, vol. 112, 02 2013.

[3] M. S.-F. Virginie Rodriguez, "Le contenu des communications relationnelles par email des enseignes : Quelle perception par le consommateur?" 2021.

[4] R. Miller and E. Charles, "A psychological based analysis of marketing email subject lines," in 2016 Sixteenth International Conference on Advances in ICT for Emerging Regions (ICTer), 2016, pp. 58–65.

[5] B. Klimt and Y. Yang, "The Enron corpus: A new dataset for email classification research," in Machine Learning: ECML 2004, J.-F. Boulicaut, F. Esposito, F. Giannotti, and D. Pedreschi, Eds.Berlin, Heidelberg Springer Berlin Heidelberg, 2004, pp. 217–226.

[6] R. Kalitvianski, "Traitements formels et sémantiques des échanges et des documents textuels liés à des activités collaboratives," Theses, Université Grenoble Alpes, Mar. 2018. [Online]. Available: https://tel.archives-ouvertes.fr/tel-01893348

[7] H. Guenoune, K. Cousot, M. Lafourcade, M. Mekaoui, and C. Lopez, "A dataset for anaphora analysis in French emails," in Proceedings of the Third Workshop on Computational Models of Reference, Anaphora and Coreference. Barcelona, Spain (online): Association for Computational Linguistics, Dec. 2020, pp. 165–175. [Online].Available: https://aclanthology.org/2020.crac-1.17

[8] A. Seyeditabari, N. Tabari, and W. Zadrozny, "Emotion detection in text: a review," 2018.

[9] P.Ekman, Basic Emotions.JohnWiley&Sons,Ltd,1999, ch.3, pp.45–60.[Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1002/0470013494.ch3

[10] A. Abdaoui, J. Az´e, S. Bringay, and P. Poncelet, "FEEL: a French Expanded Emotion Lexicon", Language Resources and Evaluation, vol. 51, no. 3, pp. 833–855, Sep. 2017. [Online]. Available: https://hal-lirmm.ccsd.cnrs.fr/lirmm-01348016

[11] A. Kumar, "An empirical examination of the effects of design elements of email newsletters on consumers' email responses and their purchase, "Journal of Retailing and Consumer Services,vol.58,p.102349,2021.[Online]. Available: https://www.sciencedirect.com/science/article/pii/S0969698920313576

[12] A. Bonfrer and X. Drèze, "Real-time evaluation of e-mail campaign performance", Marketing Science, 2009.

[13] U. Yaqub, S. A. Chun, V. Atluri, and J. Vaidya, "Analysis of political discourse on twitter in the context of the 2016 us presidential elections, "Government Information Quarterly,vol.34,no.4,pp.613–626,2017.[Online]. Available: https://www.sciencedirect.com/science/article/pii/S0740624X17301910

[14] L. Kaufman and P. Rousseeuw, Finding groups in data: an introduction to cluster analysis. John Wiley & Sons., 1990, john Wiley & Sons, New York.

[15] A. Mueller and S. Guido, Machine learning avec Python. O'Reilly Media, Inc., 2018.

Thank you for your attention

www.irisa.fr @irisa_lab

Institut de Recherche en Informatique et Systèmes Aléatoires