DEEP LEARNING FOR SENTIMENT ANALYSIS OF SMALL AND IMBALANCED REVIEWS IN RETAIL BANKING DOMAIN

PhD. Candidate: Irina F. RAICU

PhD. Supervisor: Prof. Univ. Dr. Ion SMEUREANU
ABSTRACT

Nowadays, digital transformation is driving the customer experience. Prospects or customers are significantly influenced in their purchase decision regarding a product or a service by reading online reviews. Building a Sentiment Analysis platform in the retail banking context from Romania in order to assist customers and prospects as well as financial institutions and banks is very challenging. Such a Sentiment Analysis system brings several benefits from 2 perspectives: From customer and prospect perspective, the customer satisfaction is improved by providing all the information regarding offers of banking products and services in a single place. From financial institutions and banks' perspective, the financial institution or bank can easily track the preferences of their customers, monitor the “sentiments” of their clients, collect their spontaneous feedback, build, and improve brand awareness.

Most recent studies are focusing on performing Sentiment Analysis based on Deep Learning methods. There are still a few outstanding issues that must be addressed when building a Sentiment Analysis system in the retail banking context using Deep Learning. The thesis proposes novel solutions to identified research issues. Due to lexical resources and datasets in Romanian, a dataset composed by retail banking reviews in Romania is created. A domain-specific word embedding for the retail banking context in Romanian is designed and implemented as an unsupervised Deep Learning application. In addition, supervised deep learning approaches for polarity classification of retail banking reviews in Romanian are designed and implemented. In the end, the Sentiment Analysis system is designed and implemented being composed of multiple components represented by the proposed solutions.

All the proposed solutions are evaluated through various performance criteria and compared with similar approaches identified in the research literature.

Keywords: sentiment analysis, brand analysis, deep learning, word embeddings, retail banking
## Contents

REZUMAT ......................................................................................................................................... Error! Bookmark not defined.

ABSTRACT ......................................................................................................................................... 1

ACKNOWLEDGEMENTS .................................................................................................................. Error! Bookmark not defined.

Chapter 1. Introduction ..................................................................................................................... Error! Bookmark not defined.

1.1. Motivation and Problem Statement ......................................................................................... Error! Bookmark not defined.

1.2. Research Objectives and Methodology ................................................................................. Error! Bookmark not defined.

1.2.1. Research Objective .............................................................................................................. Error! Bookmark not defined.

1.2.2. Research Methodology ....................................................................................................... Error! Bookmark not defined.

1.2.3. Research Project Evaluation ............................................................................................... Error! Bookmark not defined.

1.3. Thesis Structure ....................................................................................................................... Error! Bookmark not defined.

Chapter Summary ............................................................................................................................ Error! Bookmark not defined.

Chapter 2. Foundations and Perspectives of Artificial Intelligence ............................................. Error! Bookmark not defined.

2.1. Brief History of Artificial Intelligence .................................................................................... Error! Bookmark not defined.

2.2. Supervised Machine Learning ............................................................................................... Error! Bookmark not defined.

2.3. Neural Networks .................................................................................................................... Error! Bookmark not defined.

2.3.1. Neural Network Structure .................................................................................................. Error! Bookmark not defined.

2.3.2. Neural Network Training Process ..................................................................................... Error! Bookmark not defined.

2.3.3. Deep Neural Networks for Sequential Data ...................................................................... Error! Bookmark not defined.

2.4. Natural Language Processing .................................................................................................. Error! Bookmark not defined.

2.4.1. Tokenization ....................................................................................................................... Error! Bookmark not defined.

2.4.2. Stop-words Removal .......................................................................................................... Error! Bookmark not defined.

2.4.3. N-Grams .............................................................................................................................. Error! Bookmark not defined.

2.4.4. Word sense Disambiguating ............................................................................................... Error! Bookmark not defined.

2.4.5. Stemming ............................................................................................................................ Error! Bookmark not defined.

3. Textual data representation ........................................................................................................ Error! Bookmark not defined.

3.1. One hot encoding .................................................................................................................... Error! Bookmark not defined.
3.2. Co-occurrence representation

3.3. Term Frequency-Inverse Document Frequency

3.4. Word Embeddings

Chapter Summary

Chapter 3. Evolution of Sentiment Analysis approaches

3.1. Sentiment Analysis and polarity classification

3.2. Approaches of Sentiment Analysis

3.3. Sentiment Analysis Applicability

3.3.1. Sentiment analysis in financial domain

3.3.2. Sentiment analysis in retail banking domain

Chapter Summary

Chapter 4. Retail Banking Dataset Creation

4.1. Exploiting available textual data for retail banking domain

4.2. Research Framework for Dataset Creation

4.2.1. Identification of Data Sources

4.2.2. Data Extraction

4.2.3. Data Storage

4.2.4. Data Exploration

Chapter Summary

Chapter 5. RoBankingFastVec - Romanian Word Embeddings for Retail Banking Domain

5.1. Word2Vec Word Embeddings

5.1.1. CBOW

5.1.2. Skip-gram

5.1.3. Comparison between Skip-gram and CBOW
Chapter 5. FastText Word Embeddings

5.2. FastText Word Embeddings

5.3. Pre-trained Word Embeddings

5.4. RoBankingFastVec Creation

5.4.1. Data Sources Identification

5.4.2. Data Sources Aggregation

5.4.3. Training process

5.4.4. Embedding Space Exploration

Chapter Summary

Chapter 6. Deep Learning Experiments for Sentiment and Brand Analysis

6.1. Imbalanced Data Problem

6.2. Performance Evaluation Metrics

6.3. Methodology for performing Deep Learning Experiments

6.4. Sentiment Analysis Experiments

6.4.1. Experimental setting

6.4.2. Architecture Design

6.4.3. Implementation

6.4.4. Evaluation

6.4.5. Results and Discussion

6.5. Brand Analysis Experiments

6.5.1. Results

6.5.2. Discussions

6.6. Comparative Analysis of Classification Models from IMDB dataset

Chapter Summary

Chapter 7. Finometer – Retail banking Sentiment Analysis platform
Motivation.......................................................... Error! Bookmark not defined.

7.1. Finometer Functional Requirements ......................... Error! Bookmark not defined.

7.2. Finometer System Architecture.............................. Error! Bookmark not defined.

7.3. Presentation Layer: Finometer Case-Study ................. Error! Bookmark not defined.

7.3.1. Environment Tool ......................................... Error! Bookmark not defined.

7.3.2. Finometer: Data Presentation and Discussion .......... Error! Bookmark not defined.

Chapter Summary ..................................................... Error! Bookmark not defined.

Chapter 8. Conclusions .............................................. Error! Bookmark not defined.


8.2. Further Research Directions ................................. Error! Bookmark not defined.

8.3. Publications and Research Activities ..................... Error! Bookmark not defined.

8.3.1. Journal Papers ............................................. Error! Bookmark not defined.

8.3.2. International Conference Papers ......................... Error! Bookmark not defined.

8.3.3. Research projects ........................................... Error! Bookmark not defined.

8.3.4. International Doctoral Stage ............................. Error! Bookmark not defined.

8.3.5. Scientific seminars – Doctoral School of Economic Informatics Error! Bookmark not defined.

8.3.6. Business Conferences ...................................... Error! Bookmark not defined.

References..................................................................... Error! Bookmark not defined.

Annexes........................................................................ Error! Bookmark not defined.

Annex A - List of figures .............................................. Error! Bookmark not defined.

Annex B - List of tables ............................................... Error! Bookmark not defined.

Annex C – An excerpt from code source of RoBankingFastVvec Word Embeddings ....... Error! Bookmark not defined.

Annex D – An excerpt from code source of Sentiment Analysis Task using FastText pre-trained word embeddings in Romanian language (Imbalanced data Case) Error! Bookmark not defined.

Annex F – Overview of Finometer system ........................ Error! Bookmark not defined.