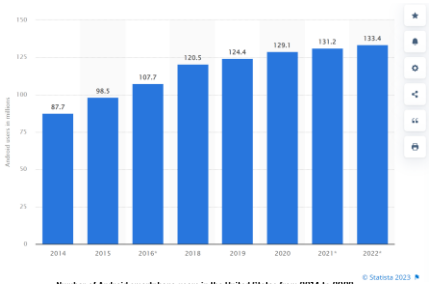
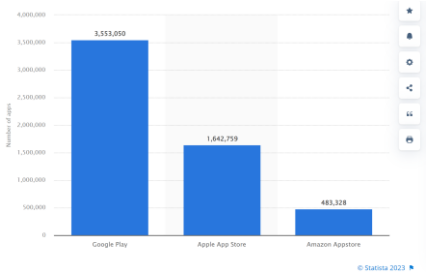
# Fine-grained Categorization of Mobile Applications through Semantic Similarity Techniques for Apps Classification

**Elena Flondor**, elena.flondor@e-uvt.ro, West University of Timișoara, Bd. Vasile Pârvan 4, Timișoara 300223
**Marc Frincu**, marc.frincu@e-uvt.ro, School of Science and Technology, Nottingham Trent University, Cliftom Campus, Nottingham, United Kingdom

## Introduction




Number of Android smartphone users in the United States from 2014 to 2022.




Number of apps available in leading app stores as of 3rd quarter of 2022

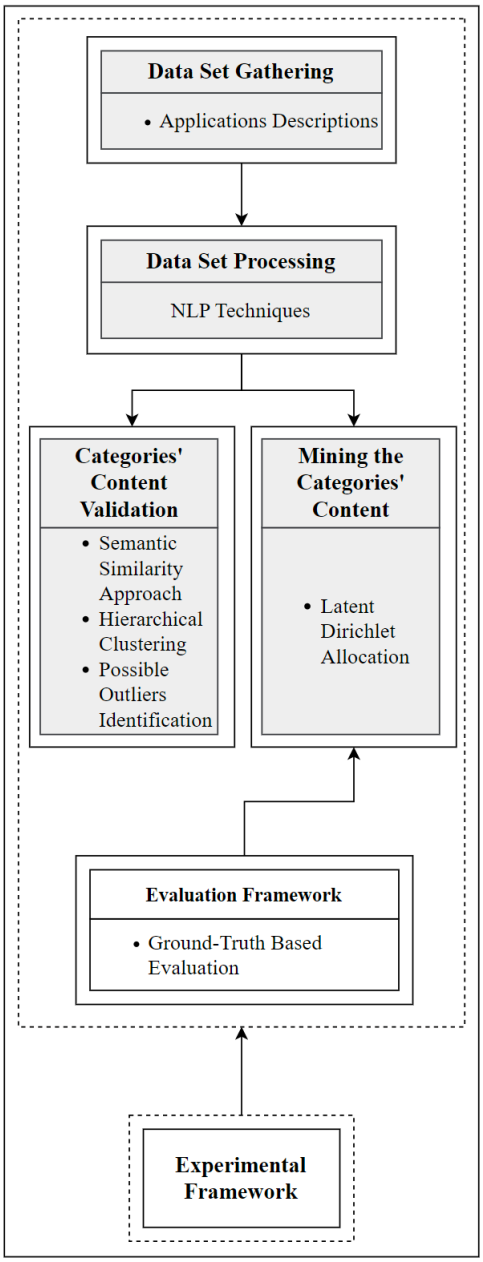
- The ongoing evolution of mobile devices in recent years has significantly impacted our lives.
- On dedicated platforms (Google Play Store, Apple App Store), developers must select a relevant category to help users find suitable apps easily.
- With thousands of apps in each category, **finding apps that match consumer interests can be challenging. Some may be misclassified**, making their discovery even more difficult.
- We focus on **validating categories' content** and **automating the process of identification of misclassified apps** for prevention, and finding possible subcategories of apps on the markets to facilitate appropriate user app selection.
- We apply **semantic similarity** [1] and **hierarchical clustering** [2] to determine misclassifications, and **Latent Dirichlet Allocation** (LDA) [3] to find apps subcategories.

| Aims | | |
|---|---|---|
| Propose an **automated method for detection of miscategorized apps.** | Discover **hidden subcategories** of apps by applying LDA on a well-defined and processed data set containing apps descriptions. | Apply LDA to **classify apps in proposed subcategories.** |

## Literature Survey

- Existing works are focused on exploiting the content of existing categories through LDA [4, 5]. A greedy hierarchical clustering was applied on features extracted from apps descriptions [6, 7]. In other work, apps descriptions were encoded using different word embeddings [8].
- As Natural Language Processing (NLP) techniques, existing works mention: stop words removal, tokenization, part of speech tagging, stemming or lemmatization [5, 8, 9].
- **Compared to the preceding work,** we **validate the content of the categories** based on the similarity between the apps descriptions and the markets recommendations and we **combine various NLP techniques**.

## Methodology




## Experiments & Results

### Outliers Identification Analysis

- Two categories of applications from the used data set presented outliers: *Medical* & *Weather*.
- **Medical**: 2 clusters - 370 applications in cluster 1 and 1 application in cluster 2. The Medical's outlier could better fit in *Education* category.
- **Weather:** 2 clusters – 293 applications in the first cluster and 1 in cluster 2. The Weather's outlier could better fit in *Game* category.

### Google Play Store – Fine-Grained Categorization

- **We removed possible outliers.**
- We **applied LDA** to explore the fine-grained categorization.
- We chose the **highest value of the coherence score** as the indicator of the best performance of LDA.
- The number of obtained **topics** sums up to **236**. It hints the existence of **more than 200 possible subcategories**.
- Averaged coherence score: **0.48**.

### App Store – Fine-Grained Categorization

- **300 educational apps** manually labeled into: *Skill-based* apps, *Content-based* apps, *Functional-based* apps, *Games* and *Misfits* [8]
- We evaluated LDA into classification in all five types of apps (Scenario 1) and into *Skill-based, Content-based, Function-based,* and *Games* – to analyze the effect of removing misclassified apps (Scenario 2).
- **Results**:

| Evaluation Metrices | Scenario 1 | Scenario 2 |
|---|---|---|
| **Precision** | 0.36 | 0.61 |
| **Recall** | 0.38 | 0.58 |
| **F2-measure** | 0.38 | 0.58 |

- Our results show that removing misclassified apps from the categories can increase the performance of a classifier by approximately 25%.
- For **Misfits** we applied the **bidirectional semantic similarity measure** (Sec. 3.3) between each app description and App Store categories recommendations [8] to discover their most suitable main categories. E.g., we found out that: *A&A Days* is more suitable for *Productivity*, *Vuga Conf* should correspond to *Social Networking*, *Weather Saying* to *Weather*, *AR Paintings* to *Graphics & Design*, etc.

## Conclusion & Future Work

### Conclusion

- We proposed the identification of misclassified mobile apps in the markets and evaluated the performance of LDA in discovering subcategories of apps.
- The experimental setup was based on averaged semantic similarity and hierarchical clustering to determine misclassified apps.
- Using apps from App Store, we showed that removing misclassified apps from the proposed data set, LDA can substantially improve classification process achieving a precision of 0.614.

### Future Work

- For future work, a closer examination of the formed LDA topics is necessary to observe the fine-grained categorization of apps published on the markets. This could suggest appropriate subcategories of apps, and markets can use this
- to improve existing classification recommendations to avoid misclassification.

## References

[1] Corley, C., Mihalcea, R.: Measuring the semantic similarity of texts. In: Proceedings of the ACL Workshop on Empirical Modeling of Semantic Equivalence and Entailment. pp. 13–18. Association for Computational Linguistics, Ann Arbor, Michigan (Jun 2005), https://aclanthology.org/W05-1203
[2] Mullner, D.: Modern hierarchical, agglomerative clustering algorithms (2011)
[3] Blei, D.M., Ng, A.Y., Jordan, M.I.: Latent dirichlet allocation. J. Mach. Learn. Res. p. 993–1022 (mar 2003)
[4] Mokarizadeh, S., Rahman, M., Matskin, M.: Mining and analysis of apps in google play pp.527–535(01 2013)
[5] Vakulenko, S., Muller, O., vom Brocke, J.: Enriching itunes app store categories via topic modeling (2014)
[6] Al-Subaihin, A.A., Sarro, F., Black, S., Capra, L., Harman, M., Jia, Y., Zhang, Y.: Clustering mobile apps based on mined textual features. In: Proceedings of the 10th ACM/IEEE International Symposium on Empirical Software Engineering and Measurement. ESEM '16, Association for Computing Machinery, New York, NY, USA (2016). https://doi.org/10.1145/2961111.2962600
[7] Harman, M., Jia, Y., Zhang, Y.: App store mining and analysis: Msr for app stores. In: 2012 9th IEEE Working Conference on Mining Software Repositories (MSR). pp. 108–111 (2012). https://doi.org/10.1109/MSR.2012.6224306
[8] Ebrahimi, F., Tushev, M., Mahmoud, A.: Classifying mobile applications using word embeddings. ACM Transactions on Software Engineering and Methodology, 31 (09 2021). https://doi.org/10.1145/3474827
[9] Lavid Ben Lulu, D., Kuflik, T.: Functionality-based clustering using short textual description: Helping users to find apps installed on their mobile device. In: Proceedings of the 2013 International Conference on Intelligent User Interfaces. p. 297–306. IUI '13, Association for Computing Machinery, New York, NY, USA (2013). https://doi.org/10.1145/2449396.2449434