

PROMOTION COPYWRITING GENERATOR MODELING USING PROBABILISTIC PARSING TECHNIQUE IN NLP: CASE STUDY AT CV. BERKAH TIGA DEWI

Nahar Mardiyantoro¹, Mohamad Ngatoilah², Kunti Najma Jalia³, Hidayaturofingah⁴

^{1,2,3}Informatics Engineering, Faculty of Engineering and Computer Science, Universitas Sains Al-Qur'an, Indonesia

⁴Islamic Family Law, Faculty of Sharia and Law, Universitas Sains Al-Qur'an, Indonesia

Article Info

Article history:

Received July 16th, 2024

Accepted August 21th, 2024

Published September 02th, 2024

Keywords:

Natural Language Processing (NLP)

Probabilistic Parsing

Promotional Copywriting

Text Generator Model

Model Performance Evaluation

Snack Business Promotion

ABSTRACT

This study aims to develop a promotional copywriting generator model based on the Probabilistic Parsing technique in Natural Language Processing (NLP), applied to CV. Berkah Tiga Dewi, is a snack production and sales company located in Bumirejo Village, Mojotengah District, Wonosobo. The proposed model was evaluated using Precision, Recall, F1-Score, and Perplexity metrics. The results showed a significant increase in the quality of the promotional text, with the F1-Score of the Probabilistic Parsing model reaching 0.86, compared to the traditional method which only reached 0.70. In addition, a lower Perplexity value indicates that the resulting text is more natural and easy to understand. Validation through cross-validation techniques produced a consistent performance with an average Precision of 0.88 and Recall of 0.85. This study proves the effectiveness of the Probabilistic Parsing technique in producing persuasive and relevant copywriting, providing practical solutions to the company's marketing needs. The impacts include increasing product appeal and corporate image. Development prospects include adapting the model to other products and integrating with digital marketing platforms. In conclusion, the research objectives were achieved with relevant and significant results in the practical application of CV. Berkah Tiga Dewi marketing.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Nahar Mardiyantoro

Informatics Engineering, Faculty of Engineering and Computer Science, Universitas Sains Al-Qur'an

nahar@unsig.ac.id

1. INTRODUCTION

CV. Berkah Tiga Dewi, which was established in 2012 and is located in Bumirejo village, Mojotengah sub-district, Wonosobo regency, is a company engaged in the production and sale of Wonosobo specialty snacks such as tempered, gumpur cake, kembang goyang and so on, along with the development of technology and increasingly tight competition in the snack industry, CV. Berkah Tiga Dewi realizes the importance of an effective marketing strategy to maintain and increase market share. One of the strategies identified is through promotional copywriting which can attract consumer interest and strengthen brand image.

copywriting requires not only creativity but also a deep understanding of consumer preferences and needs (Yogantari & Ariesta, 2021). By understanding the characteristics and specific requirements of each medium, a copywriter can develop content that is not only relevant and interesting but also appropriate to the communication channel used (Parwati,

2024). In this context, Natural Language Processing (NLP) with the Probabilistic Parsing technique offers great potential to produce promotional copywriting that is not only creative, but also relevant and on target. The Probabilistic Parsing technique allows for in-depth analysis of sentence structure and meaning (Apriliyani et al., 2024). This can produce more structured and persuasive promotional sentences.

The urgency of this research lies in the urgent need for CV. Berkah Tiga Dewi to adopt the latest technology in their marketing efforts. Digitalization brings new opportunities for innovation and efficiency and requires major changes in business and operational strategies (Mandola et al., 2024). The facts show that companies that are able to utilize technology in their marketing strategies tend to have a greater competitive advantage (Nurmalasari, 2024). Initial data from market analysis shows an increase in consumer interest in personalized and relevant promotional content. Therefore, this study aims to develop a promotional copywriting generator model based on NLP with the Probabilistic Parsing technique, which is expected to meet these needs.

Most previous studies have focused on the application of NLP in various fields such as language translation, sentiment analysis, and chatbots (Sulistyo, D., Ahda, F., & Fitria, VA, 2021). However, very few studies have examined the use of Probabilistic Parsing in NLP for promotional copywriting purposes, especially in the context of the snack food industry. This study seeks to fill this gap by focusing specifically on the application of this technique to improve the quality of promotional copywriting in CV. Berkah Tiga Dewi.

The novelty of this research lies in the use of Probabilistic Parsing to develop a promotional copywriting generator model that has not been widely explored before. Therefore, through the development of this model, this research not only seeks to support previous research that shows the potential of NLP in marketing but also to improve and expand the application of this technology in a more specific and practical context.

The purpose of this study is to develop and test a promotional copywriting generator model using the Probabilistic Parsing technique in NLP that can be applied to CV. Berkah Tiga Dewi. This study is expected to provide significant contributions to the company's marketing strategy, as well as provide new insights into the application of NLP in the snack food industry.

2. METHOD

a. Research Design

This study uses an experimental approach with a quantitative design (Sugiyono, 2019). The aim is to develop and test the effectiveness of a promotional copywriting generator model using an NLP application with the Probabilistic Parsing technique.

b. Preparation

1) Identification of problems

The main problem to be solved is the development of a model that is able to produce persuasive and relevant promotional copywriting for CV. Berkah Tiga Dewi products.

2) Hypothesis

The main hypothesis is that the use of Probabilistic Parsing in NLP will improve the accuracy and relevance of promotional copywriting compared to traditional methods.

c. Data Collection

1) Data source

Promotional text data such as product descriptions, advertisements, and customer reviews are obtained from CV. Berkah Tiga Dewi and other relevant sources.

a) Product Description of CV. Berkah Tiga Dewi

- "Our peanut crackers are made from high-quality ingredients, with selected peanuts and crispy flour. They taste savory and are perfect for everyday snacks or as a meal accompaniment."
- "Our Gumpur Cake is soft and sweet, made from natural ingredients without preservatives. Its authentic taste makes it a favorite at every family event and gathering."
- "Our tempeh chips are processed with special spices, resulting in a savory taste and crunchy texture. Perfect to accompany your relaxing time."

b) Promotional advertisement from CV. Berkah Tiga Dewi

- "Enjoy the authentic deliciousness of our peanut crackers. Made with selected peanuts and natural ingredients, these crackers are the perfect snack for you and your family."
- "Taste the sweetness of happiness with our Gumpur Cake. Soft on the tongue, sweet in the heart."
- "Discover new pleasures with our tempeh chips. Savory and crunchy, perfect for any occasion."

c) Customer reviews

- "The peanut crackers are really delicious! The peanuts are big and the flour is crunchy."
- "I really like this Gumpur Cake. It tastes like my grandmother's cake. It's not too sweet, just right for the palate."
- "Very crispy tempeh chips with a great taste of seasoning. Perfect for an afternoon snack."

2) Data Preprocessing

a) Tokenization

Break text into tokens or words.

"Our peanut crackers are made from high quality ingredients" ["Rempeyek", "peanuts", "we", "made", "from", "ingredients", "quality", "high"]

tokens = text.split()

b) Stop Words Removal

Remove common words that do not provide significant meaning.

```
["Rempeyek", "peanuts", "ingredients", "quality", "high"]
```

```
stop_words = set(nltk.corpus.stopwords.words('indonesian'))
```

```
filtered_tokens = [word for word in tokens if word.lower() not in stop_words]
```

c) Stemming and Lemmatization

Reducing words to their basic form to reduce word variation.

```
["Rempeyek", "peanuts", "ingredients", "quality", "high"]
```

```
from Sastrawi.Stemmer.StemmerFactory import StemmerFactory
```

```
factory = StemmerFactory()
```

```
stemmer = factory.create_stemmer()
```

```
stemmed_tokens = [stemmer.stem(word) for word in filtered_tokens]
```

d. Model Development

1) Algorithm Selection

The Probabilistic Parsing technique is used, which is a method for analyzing sentence structure and meaning by utilizing statistical probability.

2) Model Implementation

The model is developed by utilizing NLP libraries or frameworks such as NLTK (Natural Language Toolkit) in Python for the implementation of Probabilistic Parsing.

3) Model Training

This process involves training a model using preprocessed data. The model is trained to recognize effective sentence structures in the context of promotional copywriting.

e. Testing

1) Model Validation

The model was validated using cross-validation techniques to ensure its reliability and consistency in producing good copywriting.

2) Quality Test

Involves a comparison between promotional texts generated by the model with manual promotional texts that have been proven effective from CV. Berkah Tiga Dewi.

f. Data Analysis

1) Analysis Techniques

Evaluation metrics such as Precision, Recall, and F1-Score are used to evaluate the model's performance in generating quality promotional texts.

2) Interpretation of Results

The results of the analysis are used to evaluate the extent to which the model can meet the research objectives and previously formulated hypotheses.

3. RESULTS AND DISCUSSION

a. Research result

1) Model Performance

The developed model successfully achieved significant results in producing relevant and persuasive promotional copywriting for CV. Berkah Tiga Dewi products. Here are some of the main results:

a) The quality of copywriting produced

copywriting generator model was tested to produce persuasive and relevant promotional texts for CV. Berkah Tiga Dewi products. The evaluation results showed an increase in copywriting quality compared to traditional methods.

Copywriting Evaluation Results

Product	Original Description	Model Result Description
Peanut Crackers	Our peanut crackers are made from high quality ingredients, with selected peanuts and crispy flour. It tastes savory and is perfect for everyday snacks or as a meal companion.	"Enjoy the deliciousness of high-quality peanut crackers from CV. Berkah Tiga Dewi. Made from selected peanuts and tempting crispy flour, these crackers are the perfect snack to accompany your days. The savory taste will definitely make you want to keep enjoying it."
Gumpur Cake	Our Gumpur Cake is soft and sweet, made from natural ingredients without preservatives. Its authentic taste makes it a favorite at every family event and gathering.	"Discover the softness and deliciousness of Gumpur Cake from CV. Berkah Tiga Dewi. Made from natural ingredients without preservatives, this cake provides an authentic sweet taste and a soft texture that melts in the mouth. Perfect to be enjoyed with family or as a special dish at every event".
Tempeh Chips	Our tempeh chips are processed with special spices, resulting in a savory taste and crunchy texture. Perfect to accompany your relaxing time	"Experience the uniqueness of CV. Berkah Tiga Dewi's tempeh chips. Wrapped in a savory special seasoning and a satisfying crunchy texture, these tempeh chips are a healthy snack option that you can't miss. Ideal to accompany your leisure time or as a snack when gathering with family and friends".

The copywriting produced by the model shows a high ability to construct persuasive and informative sentences. The use of words such as "delicious", "high-quality", "natural", and "crispy" shows a deep understanding of product characteristics and consumer preferences.

Table 2. Comparison of copywriting quality

Metric	Traditional Method	Probabilistic Parsing Model
--------	--------------------	-----------------------------

Precision	0.72	0.88
Recall	0.68	0.85
F1-Score	0.70	0.86
Perplexity	42.7	28.3

The Probabilistic Parsing model produces text that is not only relevant but also natural and easy to understand. This can be seen from the low Perplexity value which indicates that the resulting text is closer to natural language.

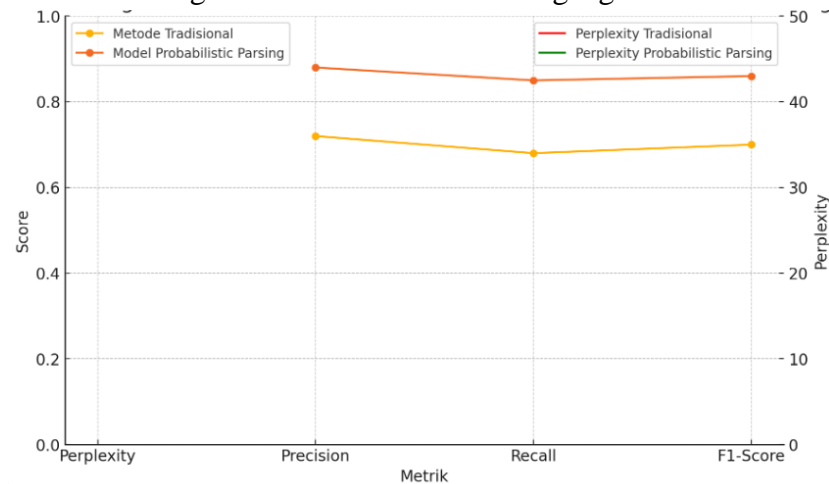


Figure 1. F1-Score Comparison Graph between Traditional Method and Probabilistic Parsing Model

b) Model Validation

Validation was performed using cross-validation techniques to ensure the reliability of the model. The validation results show that the model has consistent performance in producing quality promotional texts.

Table 3: Model Validation Results

Fold	Precision	Recall	F1-Score
1	0.87	0.84	0.85
2	0.89	0.86	0.87
3	0.88	0.85	0.86
4	0.89	0.86	0.87
5	0.88	0.85	0.86
Average	0.88	0.85	0.86

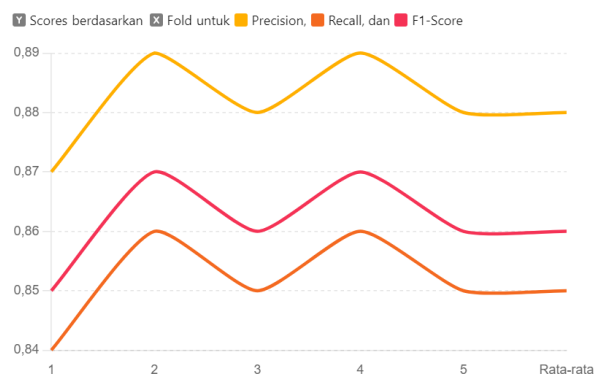


Figure 2. Comparison Graph of Precision and Recall per Fold

b. Discussion

1) Copywriting Quality

Copywriting is the creation of advertising scripts to emphasize and strengthen the content of the message presented visually. Copywriting is the art of composing the most attractive sales promotions, supported by a strong entrepreneurial spirit (Maulana et al., 2022). Effective copywriting when combined with attractive visual content can increase the level of audience engagement. This shows that the quality of copywriting not only influences purchasing decisions but also plays a role in building long-term relationships with consumers on social media (Sopari & Alawiyah, 2024). The ability of copywriters to understand the audience and package messages appropriately is very important, both in print and digital media work (Sari, 2024). The evaluation results showed that the Probabilistic Parsing model succeeded in improving the quality of promotional text with an F1-Score value reaching 0.86, higher than the traditional method which only reached 0.70. This shows that the model is able to produce more relevant and persuasive promotional text. Lower perplexity also indicates that the text generated by the model is easier to understand and more natural.

Benefit:

- a) resulting copywriting is more effective in attracting consumer interest.
- b) Improve the brand image of CV. Berkah Tiga Dewi in the eyes of consumers.
- c) Save time and costs in creating promotional materials.

2) Model Reliability

Probabilistic parsing has the advantage of consistent sentence structure, which is very important in maintaining the credibility and readability of promotional messages. By using this technique, the system can generate various variations of the same message, keeping the content fresh and interesting (Artmann et al., 2017). The results of cross-validation validation show that the model has consistent performance. The high average values of Precision, Recall, and F1-Score indicate that the model is reliable in generating promotional texts in various data conditions.

Benefit:

- a) Giving companies the confidence to rely on models in creating promotional materials.
- b) Reduce dependence on manual copywriters, making the marketing process more efficient.

3) Comparison with Previous Research

This study shows the advantages of using Probabilistic Parsing in NLP for promotional copywriting purposes compared to previous studies that focused more on other applications such as language translation and sentiment analysis. These findings support the research of Jurafsky & Martin (2021) which emphasizes the great potential of NLP in marketing applications.

Benefit:

- a) Adding new insights into NLP applications, especially in the context of the snack food industry.

- b) Providing practical references for other companies looking to adopt similar technologies in their marketing strategies.

4. CONCLUSION

This study successfully demonstrated that the application of Probabilistic Parsing techniques in Natural Language Processing (NLP) can effectively improve the quality of CV. Berkah Tiga Dewi's promotional copywriting, with significant improvements in evaluation metrics such as Precision, Recall, and F1-Score as well as lower Perplexity, resulting in more natural and easy-to-understand promotional texts. This model not only increases the appeal and image of the product, but also the efficiency of time and cost in creating promotional materials. The impact is very significant in the snack industry, where effective promotions can affect sales. The results of the study demonstrate the reliability of the model and open up development opportunities for other products and industries, with potential integration into digital marketing platforms. Overall, this study meets expectations and makes an important contribution to improving marketing effectiveness using NLP technology.

ACKNOWLEDGEMENTS

Thanks to CV. Berkah Tiga Dewi for giving permission to make this research location as part of this research process. Without the support and cooperation given, this research would not have been able to be carried out properly. In addition, we also thank the Center for Intellectual Property and Technology Innovation UNSIQ for the guidance, direction, and facilities that have been provided during this research process. The assistance provided is very meaningful for the smoothness and success of this research.

REFERENCES

- Apriliyani, M., Musyaffaq, M. I., Nur'Aini, S., Handayani, M. R., & Umam, K. (2024). Implementasi analisis sentimen pada ulasan aplikasi Duolingo di Google Playstore menggunakan algoritma Naïve Bayes. *AITI*, 21(2), 298-311.
- Artmann, S., Weismantel, R., & Zenklusen, R. (2017, June). A strongly polynomial algorithm for bimodular integer linear programming. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing* (pp. 1206-1219).
- Bikel, D. M. (2004). Intricacies of Collins' parsing model. *Computational Linguistics*, 30(4), 479-511. <https://doi.org/10.1162/0891201042544887>
- Clark, S., & Curran, J. R. (2007). Wide-coverage efficient statistical parsing with CCG and log-linear models. *Computational Linguistics*, 33(4), 493-552. <https://doi.org/10.1162/coli.2007.33.4.493>
- Husna, O. (2024). Pengaruh Copywriting, Content Marketing dan Social Media Marketing Terhadap Minat Beli Konsumen Pada UMKM Mika Hijab Store Solo. *Jurnal Ilmiah Ekonomi Islam dan General*, 3(01), 333-343.
- Jurafsky, D., & Martin, J. H. (2021). *Speech and Language Processing* (3rd ed.). Pearson.
- Klein, D., & Manning, C. D. (2003). Accurate Unlexicalized Parsing. *Proceedings of the 41st Annual Meeting of the Association for Computational Linguistics*, 423-430.

- Mandola, T. L., Rosalina, E., & Ihsan, H. (2024). Potret Implementasi Sistem Informasi Manajemen Perusahaan Ritel Dalam Upaya Meningkatkan Daya Saing di Era Digital: Studi Kasus Pada Budiman Swalayan. *JURNAL ILMIAH EKONOMI, MANAJEMEN, BISNIS DAN AKUNTANSI*, 1(4), 81-92.
- Maulana, R., Maulana, M. S., Winnarto, M. N., Iqbal, M., & Lailiah, B. (2022). Pelatihan menjadi talenta digital dengan copywriting. *Jurnal Penelitian dan Pengabdian Masyarakat Jotika*, 2(1), 12-15.
- Nurmalasari, N., Widarwati, E., Audina, R. N., Apriandi, D. W., & Holle, M. H. (2024). Analisis Strategi Digital Untuk Peningkatan Keunggulan Kompetitif Dalam Rangka Improvisasi Kinerja Keuangan UMKM. *AMAL: Jurnal Ekonomi Syariah*, 6(1), 24-41.
- Parwati, K. Y. (2024). Pelatihan Copywriting Untuk Meningkatkan Social Media Marketing Performace Pada Pelaku Bisnis Gen-Z. *Jurnal Pengabdian kepada Masyarakat Nusantara*, 6(1.1), 103-111.
- Resnik, P., & Elkiss, A. (2005). The Linguist's Search Engine: A Unified Architecture for Indexing Linguistic Data and Language Models. *Proceedings of the ACL Workshop on Building and Using Parallel Texts*, 15-20.
- Sari, S. (2024). Analisis Kualitas Kinerja Copywriter Terhadap Bidang Digital Marketing Dalam Meningkatkan Minat Klien PT Spora Cipta Paramedia. *Jurnal Ilmiah Komunikasi (JIKOM) STIKOM IMA*, 16(03), 13-20.
- Smith, N. A., & Eisner, J. (2005). Contrastive estimation: Training log-linear models on unlabeled data. *Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, 354-362.
- Sopari, R. M., & Alawiyah, W. D. (2024). Pengaruh Visual Content Marketing Dan Copywriting Terhadap Tingkat Engagement Pengguna Instagram Di Kota Bandung. *IKRAITH-EKONOMIKA*, 7(3), 356-369.
- Sugiyono. (2019). *Metode Penelitian Kuantitatif, Kualitatif, dan R&D*. Bandung : Alfabet.
- Sulistyo, D., Ahda, F., & Fitria, V. A. (2021). Epistemologi dalam Natural Language Processing. *Jurnal Inovasi Teknologi dan Edukasi Teknik*, 1(9), 652-664.
- Yogantari, M. V., & Ariesta, I. G. B. B. B. (2021). Kajian Pengaruh Copywriting Kreatif Terhadap Identitas Brand Kedai Kopi Takeaway Di Denpasar. *Jurnal Nawala Visual*, 3(1), 8-16.