Journal of Software Engineering, Information and Communication Technology (SEICT)
Journal homepage: https://ejournal.upi.edu/index.php/SEICT

Review of Banjarnese Neural Machine Translation Development With Minimal Resources

Ali Muhammad[1], Novia Winda[2], Angge Firizkiansah[3], Dita Setiawan[4], Siti Herawati Fransiska Dewi[5], Imron Maulana Rizki[6], Miri Ardiansyah[7]

[1,3,4,5,6]Program Studi Teknik Informatika, Universitas Sains Indonesia, Indonesia

[2]Program Studi Pendidikan Bahasa dan Sastra Indonesia, Universitas PGRI Kalimantan, Indonesia

[7]Program Studi Sistem Informasi, Universitas Sains Indonesia, Indonesia

Correspondence: E-mail: ali.muhammad@lecturer.sains.ac.id

ABSTRACT

With the advancement of information technology, the application of machine learning in the property industry, particularly for house price prediction, has become increasingly important. Technology plays a crucial role in speeding up and enhancing the accuracy of property buying and selling processes. Therefore, the role of machine learning technology can be utilized to meet the need for improving the accuracy of house price predictions in major cities of developing countries, such as Bandung. This research aims to analyze the effectiveness of the Artificial Neural Network and Random Forest algorithms in predicting house prices in Bandung. The data used includes house sales data in Bandung, covering land area, building area, number of bedrooms, number of bathrooms, number of parking spaces, and the subdistrict location. The analysis of the algorithms is conducted by comparing the performance testing results of both algorithms using performance metrics for regression models such as Mean Absolute Error (MAE), Mean Squared Error (MSE), Root Mean Squared Error (RMSE), and R-Square (R2). Additionally, this research analyzes which data ratio among the training, validation, and test data yields the best results. The research findings indicate that the model with a data ratio of 60:20:20 produces the best performance for both algorithms. The Random Forest algorithm demonstrates superior performance with results of MAE: 0.0470; MSE: 0.0079; RMSE: 0.0888; and R2: 0.7085.

ARTICLE INFO

Article History:
Submitted/Received 23 March 2025
First Revised 08 April 2025
Accepted 25 April 2025
First Available online 01 June 2025
Publication Date 01 June 2025

Keyword:
Banjarnese language;
Minimal resources;
Neural machine translation;
Systematic literature review.

1. INTRODUCTION

The development of artificial intelligence technology, especially in the field of natural language processing (NLP), has had a significant impact on the ability of machines to translate languages. Neural Machine Translation (NMT) has become the latest approach in automatic translation systems, replacing the previously dominant statistical approach. The implementation of NMT has spread to various world languages, including local and regional languages. One of the major challenges in the development of NMT is its application to lowresource languages, namely languages that have limited digital data corpus and supporting technology, such as Banjar (Dewangan et al., 2021; Isnaeni et al., 2024). On the other hand, the preservation and revitalization of regional languages is an important concern in the context of globalization which has lack of the existence local languages (Kamariah & Hamidah, 2023). NMT has the advantage of understanding the context of a complete sentence and producing a more natural translation that is close to the quality of human translation (Tan et al., 2021; Siu, 2023). Initial efforts have been made, such as the development of a parser and part-of-speech tagger for Banjarese (Muhammad & Kamariah, 2020; Muhammad & Widyastuti, 2024), as well as a dictionary application based on statistical inference (Purnajaya et al., 2020). However, reviews of various NMT techniques applied specifically to Banjarese are still very limited.

The lack of studies that can provide an overview of the neural machine translation approach to Banjarese — Bahasa (Indonesian) is a challenge in itself. Field data shows that Banjarese texts have unique morphological and syntactic characteristics and have not been widely documented digitally (Winda & Muhammad, 2023). This condition hold up the development of accurate and effective translation technology. Previous studies have shown that NMT provides better results than statistical approaches, especially in providing translations for languages with complex structures (Dewangan et al., 2021; Das et al., 2024). In addition, experiments on regional languages such as Makassar and Sasak show that NMT can still be adapted for low-resource languages if carried out with the right approach, such as transfer learning, model fine-tuning, or large model integration (Isnaeni et al., 2024; Wardhana et al., 2024; Team, 2024).

This study aims to review the application of various neural machine translation (NMT) techniques in translating Banjarese into Indonesian. This study also provides an analysis of the advantages and disadvantages of each NMT approach used in the formation of a neural machine translation. In addition, this study will provide technical recommendations for the development of a more effective NMT-based Banjarnese-Indonesian translation system. The benefits obtained from this study are to provide a theoretical study in the development of neural machine translation for regional languages that are included in the low-resource category. This study contributes to the field of digital technology in preserving and revitalizing the Banjarese language through a technological approach. The significant impact expected from the research on neural machine translation is to elevate the existence and value of local culture through digital technology that is adaptive to the local context.

2. METHODS

This study aims to review the application of various neural machine translation (NMT) techniques in translating Banjar language into Indonesian. To achieve this goal, this study uses the Systematic Literature Review/SLR stage. The stages of SLR are as follows:

DOI: https://doi.org/10.17509/seict.v6i1.86768
p-ISSN 2774-1656 e-ISSN 2774-1699

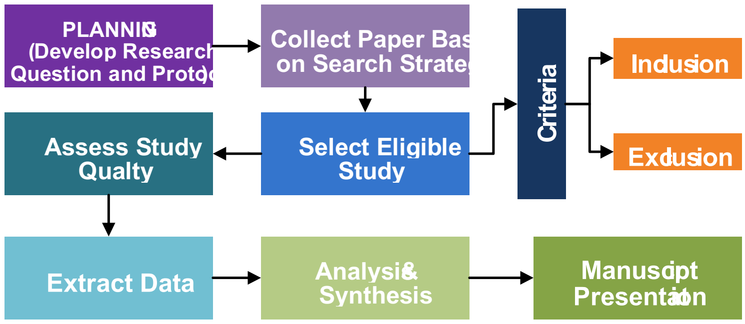

Figure 1. Stages of Conducting a Systematic Literature Review

2.1. Planning (Develop Research Question and Protocol)

The main objective of this study is to identify the latest technology in the development of neural translation machines that can be adopted into Banjarese-Indonesian. In the planning stage in the form of research questions and protocols, the research question table is in the form of a column containing questions and motivations.

2.2. Collect Paper Based on Search Strategy

At this stage, a search for papers that are related to the proposed research title is carried out. The search method is through Scopus, IEEE Explorer journals, and Google Scholar Database. The keywords used in the search are "translation machine", "neural translation machine", "Banjar language", "word dictionary".

2.3. Select Eligible Study

The learning alignment selection stage is carried out to sort between papers/journals that are in accordance with the research. This selection is based on the topic and focus of the research that is in accordance with the theme of the Banjarese-Indonesian neural translation machine and topics related to the research theme. Themes related to research in natural language processing of the Banjarese language can also be selected because they are the basis or foundation of research with related themes.

2.4. Criteria

The criteria stage is used to filter papers/journals obtained from Scopus Journal, IEEE Explorer and Google Scholar. This screening process is based on inclusion criteria and exclusion criteria. Inclusion criteria are screening criteria that determine the conditions that must be met for a study to be included in the literature review. While exclusion criteria are screening criteria by determining the conditions that cause a study to be excluded from the review even though it initially appears relevant. The criteria screening process is explained in Figure 2. While the requirements for each criterion are in **Table 2**.

DOI: https://doi.org/10.17509/seict.v6i1.86768
p-ISSN 2774-1656 e-ISSN 2774-1699

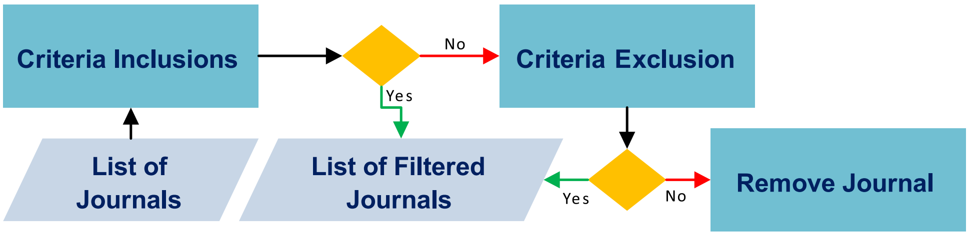

Figure 2. Article Selection Process Based on Criteria

2.5. Assess Study

The research assessment stage aims to assess the methodological quality of previous research that has been conducted. This process allows for the evaluation of the validity and reliability of research results and ensures that only quality and relevant research is used for synthesis. Aspects assessed include the suitability of the method to the objectives, availability of data or experiments, validity of results and conclusions, reproducibility of experiments, and transparency of dataset sources and codes.

2.6. Extract Data

The data extraction stage is used to obtain information that can answer the research questions outlined in Table 1. The purpose of data extraction is to collect important information systematically, prepare data for analysis and synthesis both qualitatively and quantitatively, organize important variables such as methods, datasets, results, weaknesses, and contributions, and to facilitate comparisons between studies in the final analysis stage.

2.7. Analysis & Synthesis

Analysis and synthesis stage is used to process and understand information from various studies that have been extracted. This will involve the process of identifying patterns, trends, gaps, and contradictions between the studies analyzed. Researchers will try to combine the various findings into a unified narrative according to the focus of the research. The end result of this process is to provide a strong basis for drawing conclusions, formulating recommendations, or determining the direction of further research.

2.8. Manuscript Presentation

At the manuscript presentation stage, all findings, analysis and synthesis of the research are compiled and presented systematically and academically in the form of a manuscript such as a scientific article. The purpose of this manuscript presentation is to communicate the process, results and contributions of the research to readers and the scientific community. **3. RESULTS AND DISCUSSION**

The research begins with a planning stage in the form of research questions and protocols. The research questions are outlined in Table 1 below.

Table 1. Research Question

ID	Research Question	
	Question	Motivation
RQ1	What methods are used in developing Neural Machine Translation (NMT) for regional languages?	Identifying and analyzing the appropriate renewable methods in the formation of neural machine translation (NMT).

DOI: https://doi.org/10.17509/seict.v6i1.86768
p-ISSN 2774-1656 e-ISSN 2774-1699

RQ2	What factors influence researchers in developing Neural Machine Translation (NMT)?	Identifying and analyzing the advantages and disadvantages of each method of developing automatic machine translation.
RQ3	What is the most effective approach in developing neural machine translation (NMT)?	Drawing conclusions from various methodologies on effective approaches in developing neural machine translation (NMT).

Next, a search for journals that correlate with the research is conducted. The criteria for the journals taken are included in the Inclusion criteria and Exclusion criteria in table 2 below.

Table 2. Inclusion Criteria and Exclusion Criteria

Inclusion Criteria	**Exclusion Criteria**
Journals/papers published in the last 5 years (2021-2025)	Duplicated papers/journals or journals older than 5 years
Articles are only in the form of Journals	Books, Conferences, and Long articles not related to the Research Question (RQ)
Suitability of papers/journals to the research area in the field of Information Technology and Computer Science	Papers/journals not in accordance with the research area of Information Technology and Computer Science
Research methods involve neural translation machines and Banjar language	Research methods are unclear

The results of the screening based on inclusion criteria and exclusion criteria are described in the Prisma Flow Diagram in Figure 3 below.

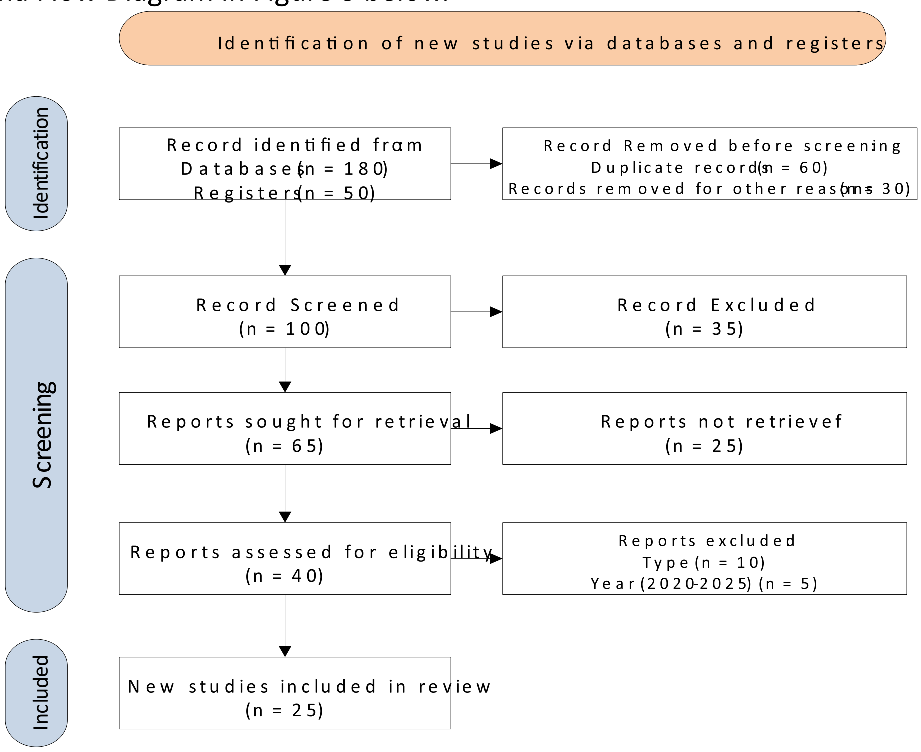

Figure 3. Literacy results after the criteria filtering process using the Prisma Flow Diagram

Next, an assessment is carried out on the articles that have been screened. The assessment

DOI: https://doi.org/10.17509/seict.v6i1.86768
p-ISSN 2774-1656 e-ISSN 2774-1699

process is a form of eligibility that the articles can be used as references in this Systematic Literature Review research. After the assessment is carried out, the articles are grouped based on the year of publication. The results of this grouping are described in **Figure 4**.

Figure 4 shows that reference research articles experienced a decline in research interest in 2022. This is understandable because in the previous year (2020 and 2021) the world was hit by Covid and forced individuals to survive. Research has increased again and the problem of regional languages must be raised as an effort to conserve and defend regional culture.

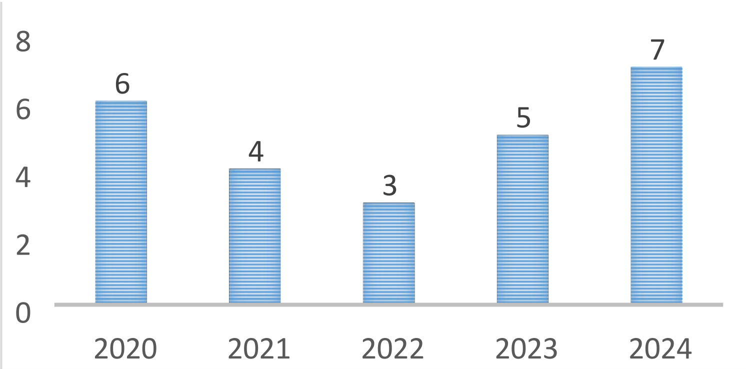

Figure 4. Findings Based on Publication Year

For further evaluation, the articles are grouped into 2 parts, namely national journals and international journals. The results of the grouping are described in Figure 5 below.

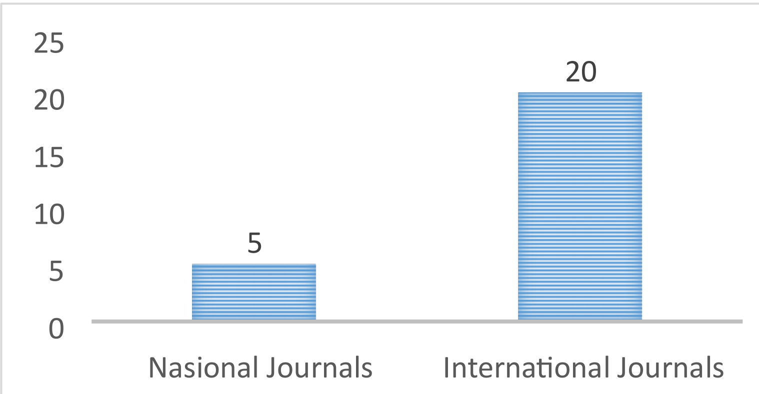

Figure 5. Results of National and International Journal Criteria

International journals discuss more strategic issues about language conservation that is on the verge of extinction due to the absorption of foreign languages. In addition, self-aware efforts to maintain the dignity of the language can be started with the conservation of regional languages through technology, especially neural translation machines. From all articles/journals obtained, data extraction is then carried out to answer the research questions that have been described in table 1 above.

RQ1: What methods are used in developing Neural Machine Translation (NMT) for regional languages?

Table 3. Neural Machine Translation Development Methods

No.	Metode	References
1	Attention Mechanisms	(Ismailia, 2023)(Lei & Li, 2023)(Tars et al., 2021)(Putri & Dewi, 2021)(Sujarwo, 2020)

DOI: https://doi.org/10.17509/seict.v6i1.86768
p-ISSN 2774-1656 e-ISSN 2774-1699

2	LLM / Transfer Learning	(Siu, 2023)(Lyu et al., 2024)(Team, 2024)
3	LSTM	(Shah et al., 2023)(Isnaeni et al., 2024)(Wardhana et al., 2024)(Tan et al., 2021)
4	Subword Segmentation	(Saunders et al., 2020)
5	*Syntactical Machine Translation* (SMT)	(Das et al., 2024)(Dewangan et al., 2021)

Research on neural machine translation is a major concern in the field of natural language processing. This research is very important to maintain the culture of regional languages that have minimal resources and are on the verge of extinction. Various methods are used to conduct analysis and development in the field of neural machine translation. Research conducted by Ismailia, Lei & Li, Tars, Putri & Dewi, and Sujarwo utilizes the existence of the Google Translate machine translation engine to build a neural machine translation engine (Ismailia, 2023; Lei & Li, 2023; Putri & Dewi, 2021; Sujarwo, 2020; Tars et al., 2021). Another study was conducted by Siu, Team and Lyu, research on the development of neural machine translation using the LLM (Large Language Model) approach, this approach allows for learning exchange and improvement for specific translation tasks (Lyu et al., 2024; Siu, 2023; Team, 2024).

Other studies use sub word segmentation. This study allows exchanging words that match the meaning contained or the context of the sentence (Saunders et al., 2020). On the other hand, neural machine translation research is also carried out using the Syntactical Machine Translation (SMT) method. This study allows the machine to replace words with words that have the same meaning or meaning without considering the context of the sentence. Additional methods or improvisation are needed so that sentences do not lose their context after being translated (Das et al., 2024; Dewangan et al., 2021). Finally, the neural machine translation research method uses the Long Short-Term Memory (LSTM) method. Research using the LSTM method requires a sentence corpus so that the neural machine translation can work. The translated words will be compared with the probability of the appearance of the intended word. This translation method is quite effective when there is a lack of experts (Isnaeni et al., 2024; Shah et al., 2023; Tan et al., 2021; Wardhana et al., 2024).

RQ2: What factors influence researchers in developing Neural Machine Translation (NMT)?

Das, Isnaeni et al. stated that neural machine translation research was developed to maintain the existence of languages from the small number of speakers of the language being studied (Das et al., 2024; Isnaeni et al., 2024; KAUR, 2013; Muhammad & Kamariah, 2020; Muhammad & Widyastuti, 2024; Tars et al., 2021; Winda & Muhammad, 2023). In addition, research is also a medium to provide applications for future needs and research. The existence of neural machine translation is very important for natural language processing such as summarization machines, question-and-answer machines, linguistic secret message hiding machines and many others (Das et al., 2024; Isnaeni et al., 2024; Purnajaya et al., 2020; Putri & Dewi, 2021; Shah et al., 2023; Widiatmika et al., 2018).

RQ3: What is the most effective approach in developing neural machine translation (NMT)?

The most applicable method to build a simple neural translation machine is to use Syntactical Machine Translation (SMT). This method is considered the most effective if there

DOI: https://doi.org/10.17509/seict.v6i1.86768
p-ISSN 2774-1656 e-ISSN 2774-1699

are no experts who can be asked as reviewers or respondents in the research. One of the challenges of using this method is that the translated sentences are out of context of the original sentence. There needs to be an improvement in the method to produce translation results that are in accordance with the desired sentence context (Barmawi & Muhammad, 2019; Das et al., 2024; Dewangan et al., 2021; Larasati, 2012; Muhammad & Kamariah, 2020; Muhammad & Widyastuti, 2024; Widiatmika et al., 2018; Winda & Muhammad, 2023).

The results of neural machine translation are measured using the Bilingual Evaluation Understudy (BLEU) application. BLEU provides a numerical score to measure how similar the machine translation results are to one or more human reference translations. The assessment method is by using N-gram Precision, Brevity Penalty and Final Score. (Barmawi & Muhammad, 2019; Briggs, 2018; Das et al., 2024; Dewangan et al., 2021; Isnaeni et al., 2024; Okta et al., 2012; Sakre, 2019; Saunders et al., 2020; Tars et al., 2021; Tattar et al., 2022).

Another method to measure the results of neural machine translation is to use the Metric for Evaluation of Translation with Explicit ORdering (METEOR). METEOR uses an automatic evaluation metric designed to assess the quality of machine translation in a way that is closer to human judgment than metrics such as BLEU. METEOR is widely used in the context of Neural Machine Translation (because it considers semantic aspects and language flexibility (Barmawi & Muhammad, 2019; Das et al., 2024; Hutchins, 1995; Sakre, 2019).

Finally, the results of the neural machine translation research were measured using the Rank-based Intuitive Bilingual Evaluation Score (RIBES). RIBES is an automatic evaluation metric used in the field of neural masschine translation to assess the quality of machine translation based on the suitability of the word order between the translation results and human references. Ribes works using the principle of word matching between languages, rank correlation coefficients and simple precision (Das et al., 2024). The testing methods used can be seen in table 4 below.

Table 4. Neural Machine Translation Results Testing Method

No.	Testing Methode	References
1	BLEU	(Barmawi & Muhammad, 2019; Briggs, 2018; Das et al., 2024; Dewangan et al., 2021; Isnaeni et al., 2024; Okta et al., 2012; Sakre, 2019; Saunders et al., 2020; Tars et al., 2021; Tattar et al., 2022)
2	METEOR	(Barmawi & Muhammad, 2019; Das et al., 2024; Hutchins, 1995; Sakre, 2019)
3	RIBES	(Das et al., 2024)

4. CONCLUSION

The development of neural machine translation can use the Attention Mechanism, Large Language Model (LLM)/Transfer Learning, Long Short-Term Memory (LSTM), Subword Segmentation, and Syntactical Machine Translation (SMT) methods. The results of the analysis show that the simplest method in developing a neural machine translation is to use the Syntactical Machine Translation (SMT) method. This method works by replacing words per word first without evaluating the sentence context. To produce a translation result that is in accordance with the sentence context, improvements to the method or the addition of a sentence context evaluation method are needed. In general, the development of neural

DOI: https://doi.org/10.17509/seict.v6i1.86768
p-ISSN 2774-1656 e-ISSN 2774-1699

machine translation is developed based on the awareness of the importance of language preservation and to prevent language extinction. The development of neural machine translation is also developed for future needs, for example for translating ancient literary works, summarizing machines, automatic question-and-answer machines, inserting secret messages linguistically, etc. The results of the neural machine translation can be evaluated using the BLEU, METEOR, or RIBES applications. The Banjar language neural machine translation should be developed as a language conservation tool in maintaining the Banjar Language and Literature.

5. REFERENCES

Barmawi, A. M., and Muhammad, A. 2019. *Paraphrasing method based on contextual synonym substitution*. *Journal of ICT Research and Applications*, *13*(3), 257–282.

Briggs, N. 2018. *Neural machine translation tools in the language learning classroom : Students ' use , perceptions , and analyses*. *The Jaltcalljournal*, *14*(1), 3–24.

Das, S. B., Panda, D., and Mishra, T. K. 2024. *Statistical Machine Translation for Indic Languages. Natural Language Processing Applications for Low-Resource Languages*, *31*(2), 328–345.

Dewangan, S., Alva, S., Joshi, N., and Bhattacharyya, P. 2021. *Experience of neural machine translation between Indian languages*. *Machine Translation*, *35*(1), 71–99.

Hutchins, W. J. 1995. *MACHINE TRANSLATION: A BRIEF HISTORY*. *Oxford: Pergamon Press*, 431–445.

Ismailia, T. 2023. *The Analysis of Machine Translation Performance on Translating Informative Text from English into Indonesian*. *Guage Teaching, Linguistics, and Literatur. Vol 3*(2), 129–138.

Isnaeni, D., Afra, N., Uliniansyah, M. T., Latief, A. D., Agency, I., Wabula, Y., and Agency, I. 2024. *Neural Machine Translation for Low-Resource Languages : Experiments on MakassarIndonesian*. *2024 International Conference on Computer, Control, Informatics and Its Applications (IC3INA) Neural*, *December*.

KAUR, S. R. E.-S. A. L. M. P. S. R. S. M. A. R. 2013. *Neural Machine Translation for Low-Resource Languages: A Survey*. *ACM Computing Surveys*, *55*(11), 1–37.

Larasati, S. D. 2012. *IDENTIC corpus: Morphologically enriched Indonesian-english parallel corpus*. *Proceedings of the 8th International Conference on Language Resources and Evaluation, LREC 2012*, 902–906.

Lei, S., and Li, Y. 2023. *English Machine translation System Based on Neural Network Algorithm Algorithm*. *3rd International Conference on Machine Learning and Big Data Analytics for IoT Security and Privacy*, *228*, 409–420.

Lyu, C., Du, Z., Xu, J., Duan, Y., Wu, M., Lynn, T., Aji, A. F., Wong, D. F., Liu, S., and Wang, L. 2024. *A Paradigm Shift : The Future of Machine Translation Lies with Large Language Models*. *ELRA Language Resource Association: CC BY-NC 4.0*, 1339–1352.

Muhammad, A., and Kamariah. 2020. *Pengurai Kalimat Bahasa Banjar Dengan Menggunakan Parser Pc-Patr*. *Jurnal Linguistik Komputasional (JLK)*, *3*(1), 20.

Muhammad, A., dan Widyastuti, N. 2024. *Pengembangan Aplikasi Part-of-Speech Tagger Bahasa Banjar Menggunakan Metode Pengembangan DevOps*. *JIKOMTI: Jurnal Ilmiah Ilmu Komputer Dan Teknologi Informasi*, *1*(1).

DOI: https://doi.org/10.17509/seict.v6i1.86768
p-ISSN 2774-1656 e-ISSN 2774-1699

Okta, B., Miranda, S., Yuliansyah, H., and Biddinika, M. K. 2012. *Machine Translation Indonesian Bengkulu Malay Using Neural Machine Translation-LSTM*. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*.

Purnajaya, A. R., Indriani, F., dan Faisal, M. R. 2020. *Pengenalan Suara Pada Kamus Banjar - Indonesia dan Indonesia - Banjar Menggunakan Statistik Inferensi*. *Jurnal Ilmiah Informatika (JIF)*, *08*(01).

Putri, S. A., and Dewi, H. D. 2021. *The Translation Of Entertainment News From English To Indonesian With Machine Translation*. *Journal of English Language and Culture*, *11*(2), 153–168.

Sakre, M. M. 2019. *Machine translation status and its effect on business*. *Journal of the ACS*, *10*(May).

Saunders, D., Sallis, R., and Byrne, B. 2020. *Neural Machine Translation Doesn ' t Translate Gender Coreference Right Unless You Make It*. *Proceedings Ofthe Second Workshop on Gender Bias in Natural Language Processing*, 35–43.

Shah, V., Valiullah, S., and Makwana, M. Y. 2023. *Neural machine translation to local languages*. Indus University.

Siu, S. C. 2023. *Revolutionizing Translation with AI : Unravelling Neural Machine Translation and Generative Pre-trained Large Language Models An Overview of Deep Learning*. *School of Translation and Foreign Languages*, 1–19.

Sujarwo. 2020. *Students ' Perception s of Using Machine Translation Tools In the EFL Classroom*. *Al-Lisan : Jurnal Bahasa*, *5*(2), 230–241.

Tan, Z., Wang, S., Yang, Z., Chen, G., and Huang, X. 2021. N*eural machine translation : A review of methods , resources , and tools. AI Open*, *1*(October 2020), 5–21.

Tars, M., Andre, T., and Fiˇ, M. 2021. *Extremely low-resource machine translation for closely related languages*. *Proceedings of the 23rd Nordic Conference on Computational Linguistics (NoDaLiDa)*, 41–52.

Tattar, andre, Purason, T., Kuulmets, H.-A., Luhtaru, A., Ratsep, L., Tars, M., Pinnis, M., Bergmanis, T., and Fishel, M. 2022. *Open and Competitive Multilingual Neural Machine Translation in Production*. *Baltic J. Modern Computing*, *10*(3), 422–434.

Team, N. 2024. *Scaling neural machine translation to 200 languages*. *Nature*, *630*.

Wardhana, H., Dharma, I. M. Y., Marzuki, K., and Hidayatullah, I. S. 2024. *Implementation of Neural Machine Translation in Translating from Indonesian to Sasak Language*. *Matrik: Jurnal Manajemen, Teknik Informatika, Dan Rekayasa Komputer*, *23*(2), 465–476.

Widiatmika, P. W., Bagus, I., Ari, M., Made, N., and Widya, Y. 2018. *Examining The Result Of Machine Translation For Linguistic Textbook From English To Indonesian*. *The 2nd English National Seminar English Education Study Program STKIP PGRI PACITAN*, 54–65.

Winda, N., dan Muhammad, A. 2023. *Pengembangan Parsing PCPATR sebagai Preservasi Bahasa dan Sastra Banjar. Jurnal Onoma: Pendidikan, Bahasa Dan Sastra*, *9*(2), 2023.

DOI: https://doi.org/10.17509/seict.v6i1.86768
p-ISSN 2774-1656 e-ISSN 2774-1699