INTERNATIONAL UNIVERSITY OF APPLIED SCIENCES

SIGUL 2022 @ LREC 2022
The 1st Annual Meeting of the ELRA/ISCA Special Interest Group on Under-Resourced Languages

OCHILBEK RAKHMANOV AND TIM SCHLIPPE

SENTIMENT ANALYSIS FOR HAUSA: CLASSIFYING STUDENTS' COMMENTS

Marseille, France
June 25, 2022

AGENDA

1

INTERNATIONAL UNIVERSITY OF APPLIED SCIENCES

INTRODUCTION

MOTIVATION: UN Sustainable Development Goal 4

SUSTAINABLE DEVELOPMENT GOALS

Image Sources: United Nations: Sustainable Development Goals: 17 Goals to Transform our World (2021); OpenClipart-Vectors/154119/Pixabay.

MOTIVATION: UN Sustainable Development Goal 4

SUSTAINABLE DEVELOPMENT GOALS

ANALYSIS

FEEDBACK LOOP

IMPROVE EDUCATION

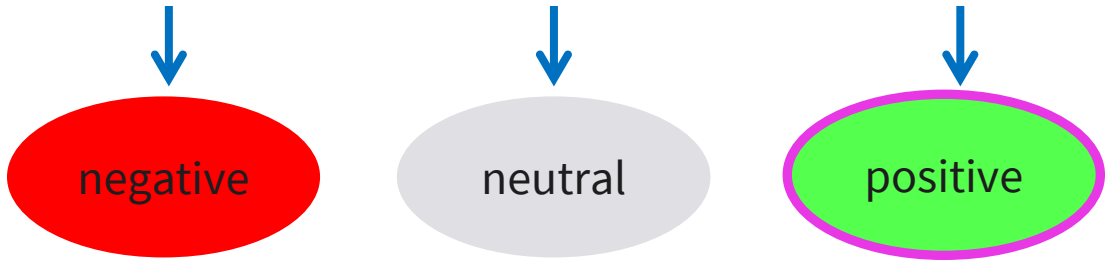
Image Sources: United Nations: Sustainable Development Goals: 17 Goals to Transform our World (2021); OpenClipart-Vectors/154119/Pixabay.

SENTIMENT ANALYSIS: Classifying student comments

INTERNATIONAL UNIVERSITY OF APPLIED SCIENCES

extracting subjective information from text such as mood,
e.g., classify student comments in course evaluations.

> *"Being a maths lecturer it's as easy as being other courses lecturer. Most students have a hard time understanding mathematics which requires constantly explaining over and over again and not all lecturers have that patience. But as for me, I have nothing much to say but he's a very good lecturer."*



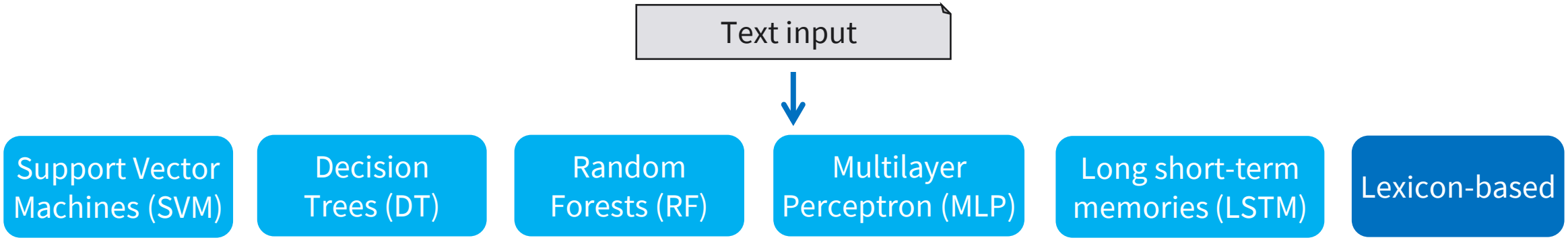
Text Source: Jay (2016), para. 1. Graphic Source: Custom Depiction.

2

INTERNATIONAL UNIVERSITY OF APPLIED SCIENCES

RELATED WORK

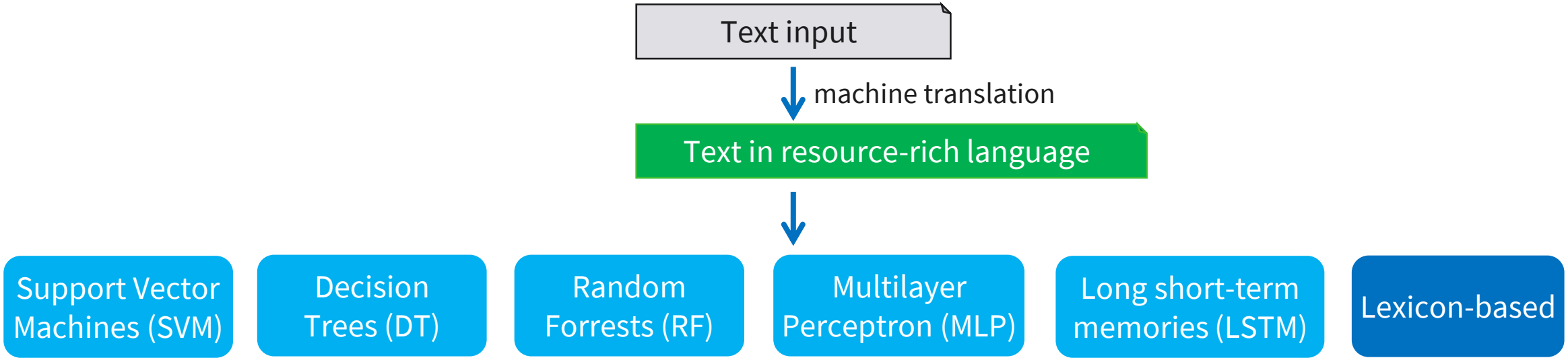
RELATED WORK: Sentiment Analysis



Balahur and Turchi (2014), Nguyen et al. (2018), Kumar and Sharan (2020), Rakhmanov (2020), ...

Kolchyna et al. (2015), Kotelnikova et al. (2021), ...

RELATED WORK: Cross-lingual Sentiment Analysis



Balahur and Turchi (2014), Lin et al. (2014),
Vilares et al. (2017), Can et al. (2018)
...

INTERNATIONAL UNIVERSITY OF APPLIED SCIENCES

HESAC

HAUSA

INTERNATIONAL UNIVERSITY OF APPLIED SCIENCES

- **under-resourced** **language**
- **lingua franca** in many countries

- **50-100 million** **speakers**, widely spoken in West Africa (Abubakar et al., 2019)
- Spoken in Nigeria, Southern Niger, Cameroon, Togo, Chad, Benin, Burkina Faso, and Ghana, etc.
- Online text resources available

HESAC

INTERNATIONAL UNIVERSITY OF APPLIED SCIENCES

The Hausa-English Sentiment Analysis Corpus For Educational Environments

- **40k comments** labeled as **positive**, **neutral**, **negative**
- **4k students** commented **524 courses** taught by **203 instructors** at Nile University of Nigeria
- To contribute to the improvement of under-resourced languages, we share the corpus with the research community:

https://github.com/MrLachin/HESAC

HESAC

INTERNATIONAL UNIVERSITY OF APPLIED SCIENCES

The Hausa-English Sentiment Analysis Corpus For Educational Environments

★★★ ✓cross-checked

“Being a maths lecturer it’s as easy as being other courses lecturer. Most students have a hard time understanding mathematics which requires constantly explaining over and over again and not all lecturers have that patience. But as for me, I have nothing much to say but he’s a very good lecturer.”

(1) machine-translated
(2) corrected
(3) cross-checked

✓

“Kasancewa malamin lissafi yana da sauki kamar kasancewa wasu darussan darussan da ke da wahalar fahimta game da ilimin lissafi wanda yake buatar bayani koyaushe da kuma hauri. Amma ni ni ba abin da zan ce sai dai shi babban malami ne.”

INTERNATIONAL UNIVERSITY OF APPLIED SCIENCES

SENTIMENT ANALYSIS FOR HAUSA

SENTIMENT ANALYSIS: Students‘ Feedback

INTERNATIONAL UNIVERSITY OF APPLIED SCIENCES

Text Input

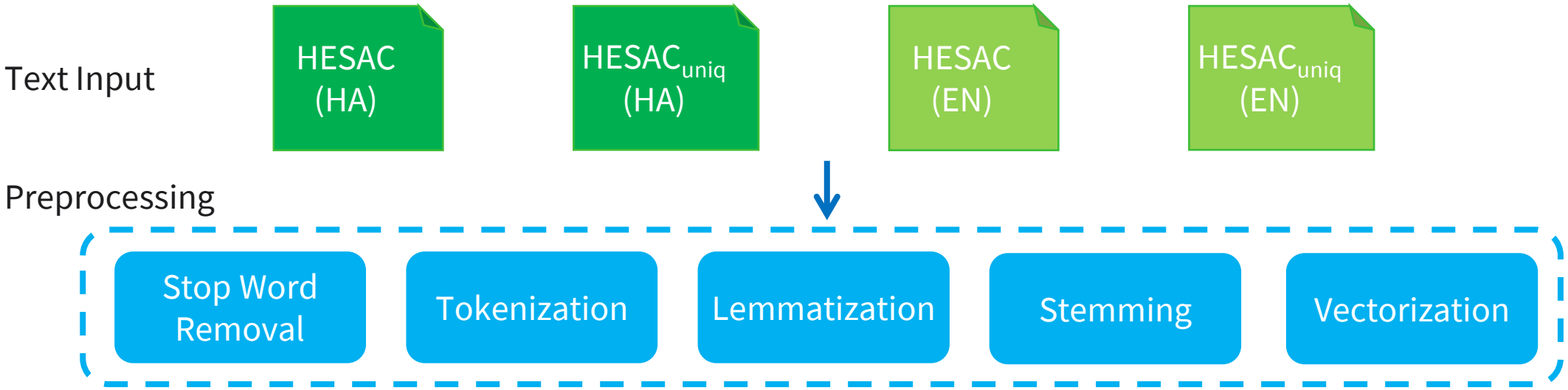
HESAC (HA)

$HESAC_{uniq}$ (HA)

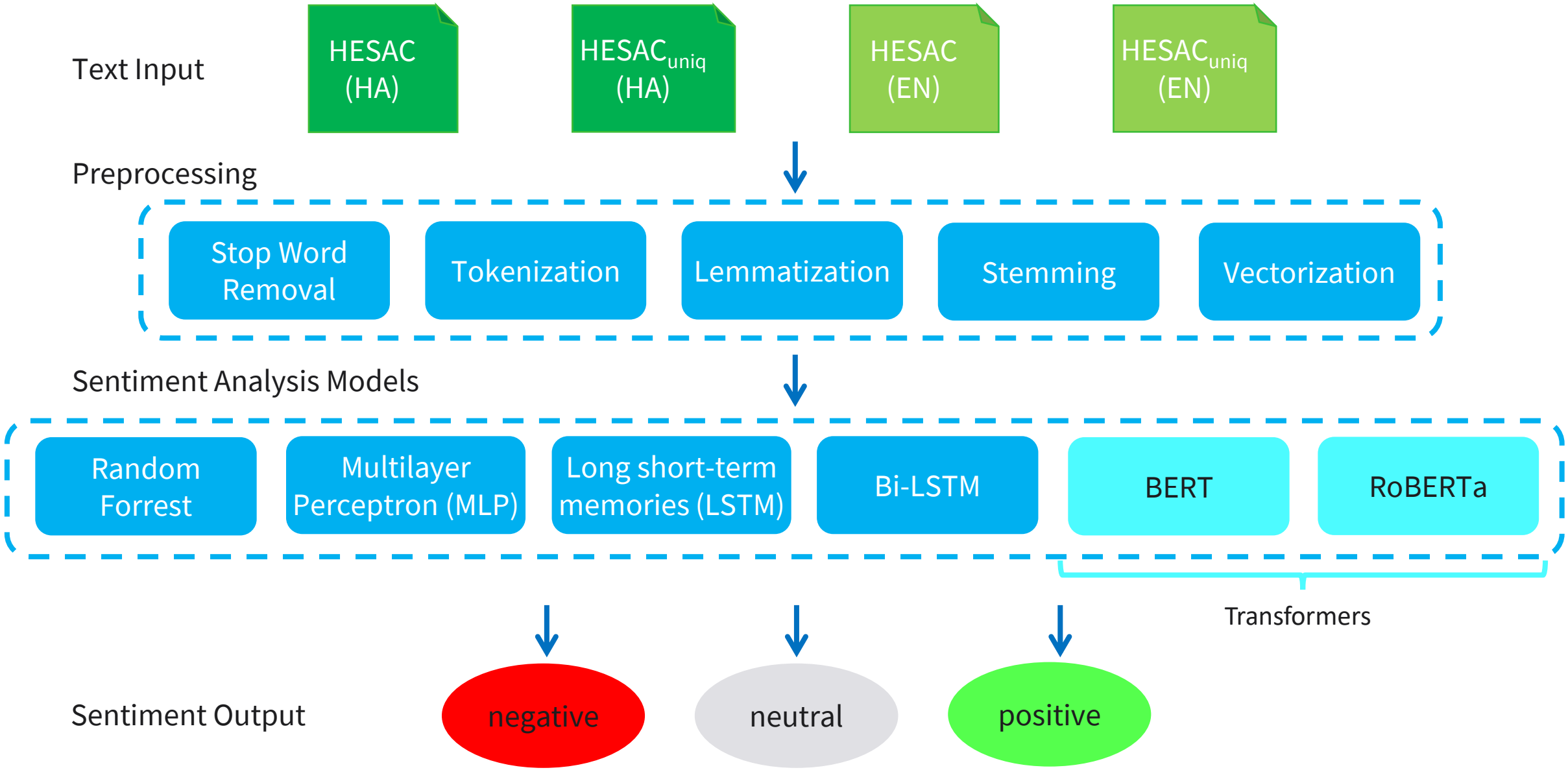
HESAC (EN)

$HESAC_{uniq}$ (EN)

SENTIMENT ANALYSIS: Students‘ Feedback



SENTIMENT ANALYSIS: Students‘ Feedback



SENTIMENT ANALYSIS: Students‘ Feedback

INTERNATIONAL UNIVERSITY OF APPLIED SCIENCES

ENGLISH

Method	training	$training_{uniq}$
RF	96.3	95.1
MLP	96.3	94.6
LSTM	97.6	94.4
Bi-LSTM	97.5	94.4
BERT	**98.7**	95.9
RoBERTa	98.5	95.3

monolingual

HAUSA

Method	training	$training_{uniq}$
RF	94.7	92.0
MLP	95.7	91.3
LSTM	96.0	92.4
Bi-LSTM	96.0	92.2
BERT	**96.9**	94,9
RoBERTa	96.4	94.5

cross-lingual

Method	training	$training_{uniq}$
RF	97.1	92.7
$RF_{stemming}$	97.3	92,8
MLP	97.0	90.8
$MLP_{stemming}$	97.1	91.1
LSTM	96.2	90.9
$LSTM_{stemming}$	**97.4**	91.4
Bi-LSTM	96.7	91.0
$Bi\text{-}LSTM_{stemming}$	97.0	91.4
RoBERTa	96.3	92.0
$RoBERTa_{stemming}$	96.3	92.0

monolingual

5

iu INTERNATIONAL UNIVERSITY OF APPLIED SCIENCES

CONCLUSION AND FUTURE WORK

CONCLUSION AND FUTURE WORK

INTERNATIONAL UNIVERSITY OF APPLIED SCIENCES

Conclusion

- Corpus of more than 40k comments—the Hausa-English Sentiment Analysis Corpus For Educational Environments (**HESAC**)
- Investigated **monolingual and cross-lingual approaches** for Hausa to classify student comments in course evaluations
- Proposed a novel **stemming algorithm** for Hausa to improve accuracy
- Experimented with **removing duplicates** from the training set, but this resulted in deterioration of the systems
- **Monolingual approaches** for Hausa slightly outperform cross-lingual ones
- Best Hausa model results in an **accuracy of 97.4% on HESAC**

CONCLUSION AND FUTURE WORK

INTERNATIONAL UNIVERSITY OF APPLIED SCIENCES

Conclusion

- Corpus of more than 40k comments—the Hausa-English Sentiment Analysis Corpus For Educational Environments (**HESAC**)
- Investigated **monolingual and cross-lingual approaches** for Hausa to classify student comments in course evaluations
- Proposed a novel **stemming algorithm** for Hausa to improve accuracy
- Experimented with **removing duplicates** from the training set, but this resulted in deterioration of the systems
- **Monolingual approaches** for Hausa slightly outperform cross-lingual ones
- Best Hausa model results in an **accuracy of 97.4% on HESAC**

Furture Work

- **Tackle performance losses** with long sentences that contain both positive and negative aspects
- **System combination**
- Add **topic identification** to extract even more valuable information from the students' feedback

INTERNATIONAL UNIVERSITY OF APPLIED SCIENCES

THANK YOU

Tim Schlippe
tim.schlippe@iu.org

REFERENCES

INTERNATIONAL UNIVERSITY OF APPLIED SCIENCES

Literature

- **United Nations: Sustainable Development Goals: 17 Goals to Transform our World (2021)**: https://www.un.org/sustainabledevelopment/sustainable-development-goals
- **Balahur, A. and Turchi, M. (2014)**: *Comparative Experiments using Supervised Learning and Machine Translation for Multilingual Sentiment Analysis*. Comput. Speech Lang., 28:56–75.
- **Nguyen, P. X. V., Hong, T. V. T., Nguyen, K. V., and Nguyen, N. L.-T. (2018)**: *Deep Learning versus Traditional Classifiers on Vietnamese Students' Feedback Corpus*. 5th NAFOSTED Conference on Information and Computer Science (NICS).
- **Kumar, A. and Sharan, A., (2020)**: *Deep Learning-Based Frameworks for Aspect-Based Sentiment Analysis*, pages 139–158. Springer Singapore.
- **Rakhmanov, O. (2020)**: *A Comparative Study on Vectorization and Classification Techniques in Sentiment Analysis to Classify Student-Lecturer Comments*. Procedia Computer Science, 178:194–204.
- **Kolchyna, O., Souza, T. T. P., Treleaven, P. C., and Aste, T. (2015)**: *Twitter Sentiment Analysis: Lexicon Method, Machine Learning Method and Their Combination*. arXiv: Computation and Language.

Images

- **Images provided by OpenClipart-Vectors/154119/Pixabay**. (https://pixabay.com/vectors/international-project-world-154119 [last access: 16.16.2021])

REFERENCES

INTERNATIONAL UNIVERSITY OF APPLIED SCIENCES

Literature

- **Kotelnikova, A., Paschenko, D., Bochenina, K., and Kotelnikov, E. (2021)**. *Lexicon-based Methods vs. BERT for Text Sentiment Analysis*. In AIST.
- **Lin, Z., Jin, X., Xu, X., Wang, Y., Tan, S., and Cheng, X. (2014)**: *Make It Possible: Multilingual Sentiment Analysis without Much Prior Knowledge*. In 2014 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT), volume 2, pages 79–86.
- **Vilares, D., Alonso Pardo, M., and Gómez-Rodríguez, C. (2017)**: *Supervised Sentiment Analysis in Multilingual Environments*. Information Processing Management, 53, 05.
- **Can, E. F., Ezen-Can, A., and Can, F. (2018)**: *Multilingual Multilingual Sentiment Analysis: An RNN-Based Framework for Limited Data*. In ACM SIGIR 2018 Workshop on Learning from Limited or Noisy Data.
- **Abubakar, A. I., Roko, A., Muhammad, A., and Saidu, I. (2019)**: *Hausa WordNet: An Electronic Lexical Resource*. Saudi Journal of Engineering and Technology, 4(8):279–285.