



AI Assisted Message Distribution Email Categorization

Raghupati Bhetwal^{1,+}, Oguzcan Sencan^{1,+}

¹⁾ University of Applied Sciences and Arts Dortmund, Department of Computer Science (FB4), 44227 Dortmund, Deutschland
^{+) Authors have contributed equally}

Introduction

- Email communication is a fundamental part of modern communication for individuals and organizations.
- Managing and categorizing incoming emails efficiently has become a challenge due to the increasing volume of messages.
- Automated email classification systems have emerged to classify emails based on content and recipients.
- This report presents the results of an email classification project focusing on predicting recipients using the Enron dataset.

Methods

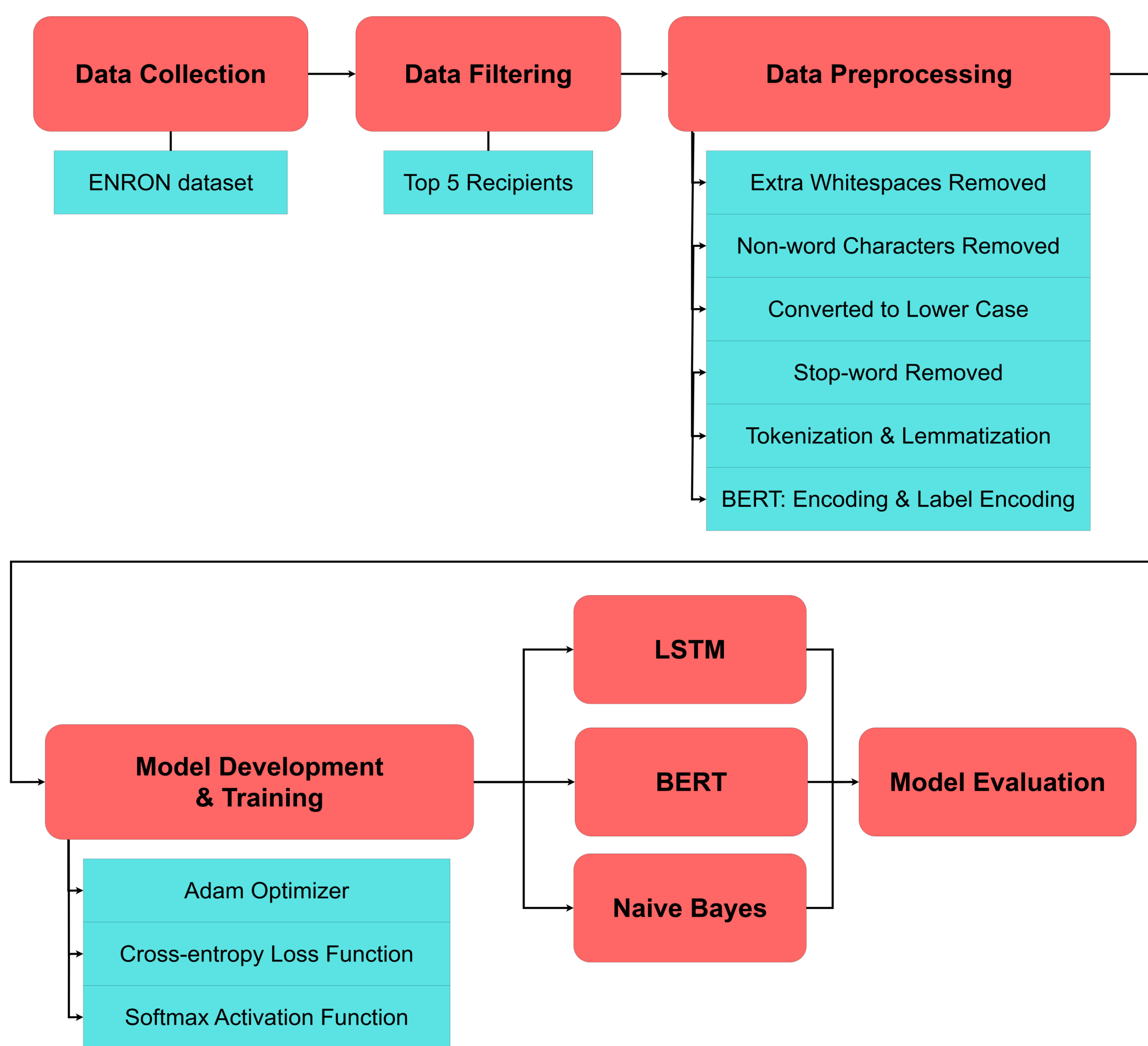


Figure : Process Flow of ML Models

Prediction

Subject: Re: Enron Response to San Diego Request for Gas Price Caps

Recognizing the time constraints you face, I've tried to 1) clear up a few inaccuracies and 2) massage some of the sharper language without taking a chainsaw to the otherwise good job.

Original Recipient (2nd email): jeff.dasovich@enron.com

Original
Email

Predicted
Recipient

```
print("Predicted Recipient (2nd email):", predicted_label_2)

[33]

... 1/1 [=====] - 0s 19ms/step
Predicted Recipient (2nd email): ['jeff.dasovich@enron.com']
```

Results

LSTM Model Training and Evaluation:

- BERT performed comparatively higher accuracy than LSTM and NB Mode.

| Model | Accuracy |
|-------|----------|
| BERT | 0.98 |
| LSTM | 0.97 |
| NB | 0.93 |

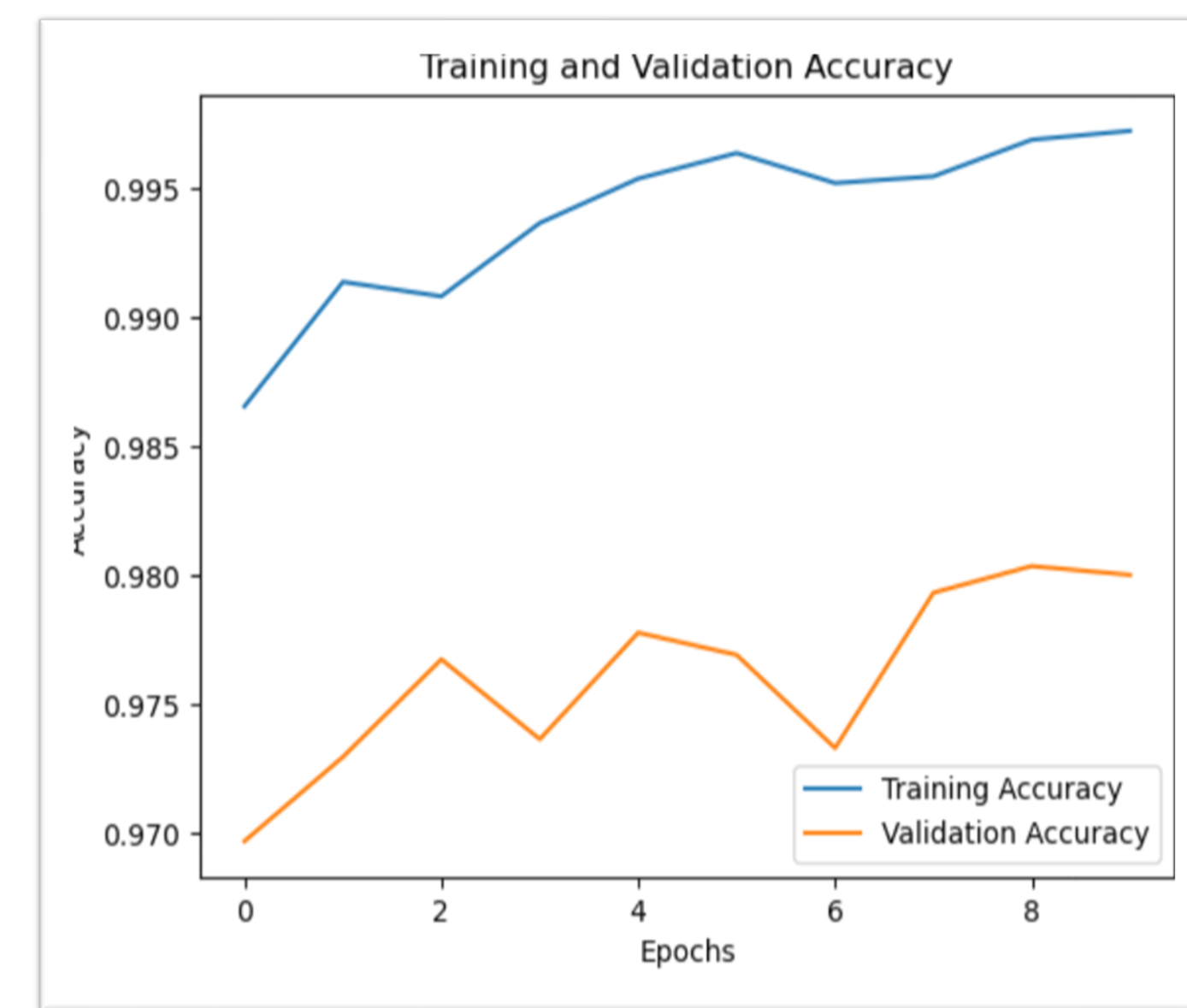
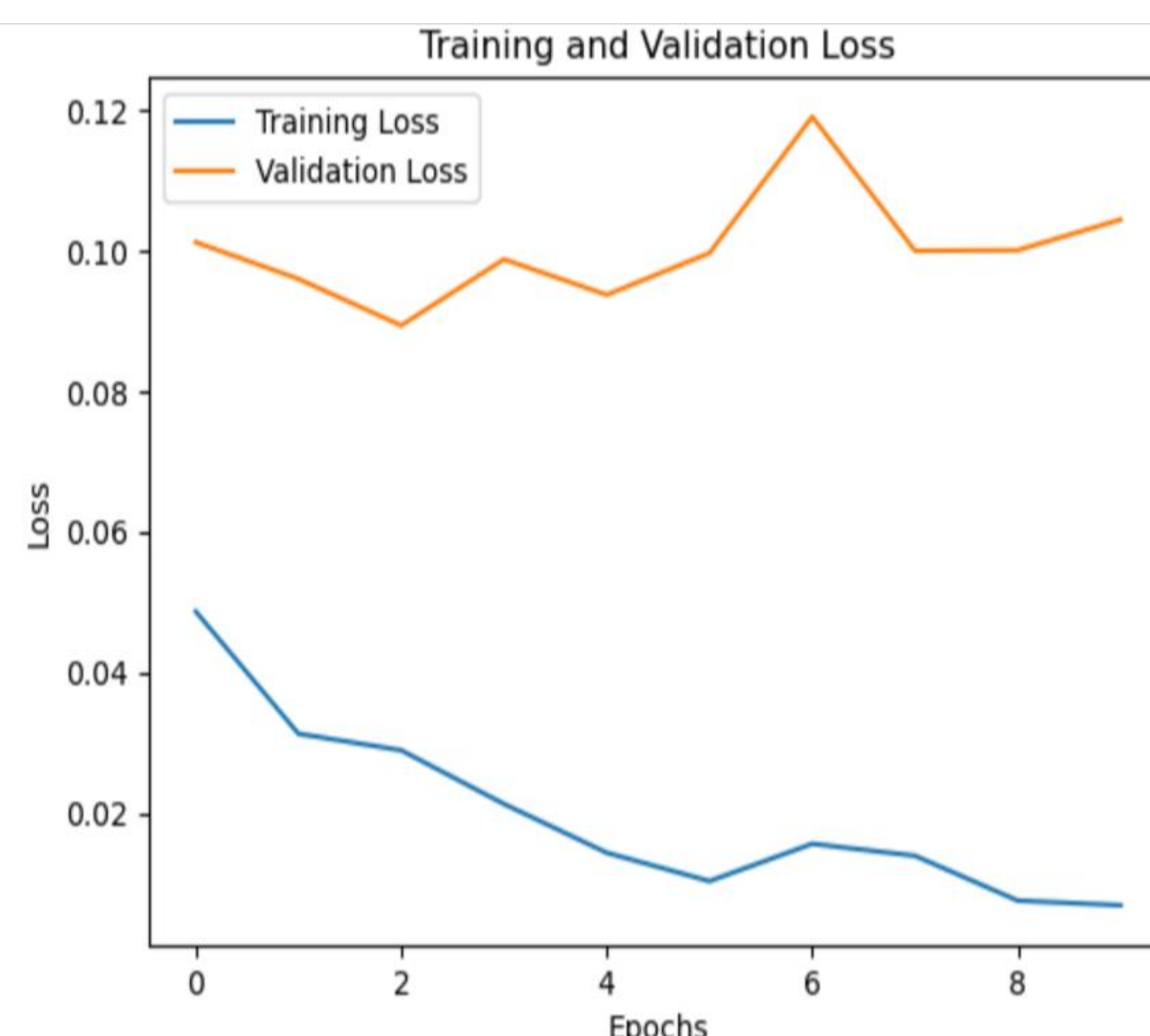


Figure : Graph representation between Training Loss/Accuracy with Validation loss/Accuracy

The model demonstrated low validation loss and high validation accuracy, indicating its ability to generalize well to unseen data.

Discussion

- The utilization of the Enron dataset and focusing on the top 5 recipients provided valuable insights into email classification based on recipient' information.
- The model's ability to capture sequential dependencies proved beneficial in accurately classifying emails.
- However, the model's performance may vary depending on the dataset and specific email classification task.
- The Enron dataset, known for its diversity and real-world nature, served as a representative scenario for this project.
- The developed email classification system can contribute to the automation of email management, enhancing productivity and efficiency.

Conclusion

- The Email classification project successfully developed an automated system for accurately classifying emails based on recipient information.
- The LSTM, BERT model, trained on pre-processed data and leveraging sequential modeling, achieved a high level of accuracy.
- The project provides insights into email communication patterns and contributes to the automation of email management.
- Future work could involve further optimization of the model, exploring alternative algorithms, and evaluating the system on different datasets.

References:

- [1] <https://www.kaggle.com/datasets/wcukierski/enron-email-dataset>
[2] https://www.tensorflow.org/api_docs/python/tf/keras/layers/LSTM

Project: This poster presentation was prepared within the DiffPro-ML project. DiffPro-ML is funded by the Federal Ministry of Education and Research under the funding code 01S17075.

Contact:

Author 1
Oguzcan Sencan001@stud.fh-
dortmund.de

Author 2
raghupati.bhetwal001@stud.fh-
dortmund.de

Christoph M. Friedrich
Emil-Figge-Straße 42
44227 Dortmund, Germany
+49 (0)231 755-6796
[christoph.friedrich@fh-
dortmund.de](mailto:christoph.friedrich@fh-dortmund.de)