

Klassifikation von geschäftlichen Emails

Studiengang: BSc in Informatik | Vertiefung: Data Engineering
Betreuer: Prof. Dr. Jürgen Vogel
Experte: Martin Arnold

In dieser Studie wird der Einsatz von Deep Learning zur automatischen Erkennung und Kategorisierung von Geschäftsposteingängen durch Textanalyse untersucht. Das Ziel der Studie besteht darin, effiziente Methoden zu entwickeln, um E-Mail-Inhalte zu identifizieren, zu kategorisieren und gezielt zu verarbeiten. Dadurch soll in Zukunft eine automatisierte Verarbeitung von E-Mails ermöglicht werden.

Ausgangslage

In der Firma Cadec AG werden Kundenanfragen oft an die Hauptmailadresse gesendet. Im Postfach sind neben E-Mails von Neukunden auch Support-Anfragen und Offerten von Stammkunden. Das HR-Team sortiert die E-Mails aus und leitet sie weiter. Mit dem Ziel der Automatisierung soll ein Deep Learning-basiertes Klassifikationsverfahren durch Textanalyse entwickelt werden. Die Motivation im Projekt ist es, das HR-Team zu entlasten und langfristig Ressourcen einzusparen.

Umsetzung

Gemeinsam mit meiner Firma wurde das Projektziel definiert, das die Kategorisierung der E-Mails umfasst. Die Anforderungen für die Kategorisierung wurden festgelegt. Anschließend wurden die Daten beschafft, indem der Outlook-Maileingang gemäß den Vorgaben annotiert wurde, um einen Goldstandard zu erstellen. Auch wurde in einem früheren Projekt bereits ein erstes Machine Learning-Modell evaluiert. Im nächsten Schritt des Bachelor Thesis erfolgte die Vorverarbeitung der E-Mail Daten, bei dem die Behandlung von Mail-Threads und die Anpassung der Vokabulargröße

durchgeführt wurden, um die Leistung des Deep Learning-Modells zu verbessern. Ausserdem wurden die Parameter des BERT-Modells optimiert

Resultate

Eine erste Evaluation des Deep Learning-Modells zeigte vielversprechende Ergebnisse bei der Vorhersage von E-Mails mit einer Genauigkeit von 83%. Allerdings ist das derzeitige Datenvolumen noch begrenzt. Um die Leistung weiter zu verbessern und automatisierte Integrationen zu ermöglichen, ist eine Erhöhung des Datenvolumens erforderlich. Aktuell ist das Modell eher für die Unterstützung bei der E-Mail-Bearbeitung geeignet, anstatt vollständig automatisiert eingesetzt zu werden.

Ausblick

Für die Zukunft besteht das Potenzial, das Modell in die Automatisierung von E-Mail-Bearbeitungsprozessen einzubinden. Durch kontinuierliche Optimierungen und Anpassungen soll die Leistung weiter gesteigert werden, um die Erkennung von E-Mails in der Firma zu verbessern und automatisierte Funktionen zu unterstützen.



Vipushan Ratnasabapathy

Guten Tag Herr Blagovic

Wir haben im September bei euch eine Antenne für Affolter Martin und Wenger Daniela bestellt.

Ist diese Antenne als Zusatz gedacht damit der Kunde 3G empfangen kann?

Danke für Ihre Rückmeldung.
Sonnige Grüsse

Leitung Elektroabteilung
Eidg. Dipl. Elektroinstallateur

T: +41 (0)27 934
M: +41 (0)79 327

Guten Tag Herr [PERSON][NEWLINE][NEWLINE][NEWLINE]

Wir haben im September bei euch eine Antenne für [PERSON] [PERSON]und [PERSON] [PERSON] bestellt. [NEWLINE]

Ist diese Antenne als Zusatz gedacht damit der Kunde 3G empfangen kann? [NEWLINE][NEWLINE]Danke für Ihre Rückmeldung.



Der eingegebene E-Mail-Text:
Guten Tag Herr [PERSON]

[NEWLINE][NEWLINE][NEWLINE][NEWLINE]

Wir haben im September bei euch eine Antenne für [PERSON] [PERSON]und [PERSON]

[PERSON] bestellt. [NEWLINE]Ist diese Antenne als Zusatz gedacht damit der Kunde 3G empfangen kann? [NEWLINE][NEWLINE]

[NEWLINE]Danke für Ihre Rückmeldung.

Das E-Mail wurde als Kategorie Support erkannt.

Support: 61.46333694458008%
Auftrag/Offerte: 21.754848957061768%
Newsletter/Spam: 16.781818568706512%

E-Mail Auswertung vom DL Modell

Vorverarbeitung der E-Mail Daten