

# Forschungspapier: KI-gestützte Dokumentenanalyse

## Forschungspapier: KI-gestützte Dokumentenanalyse

### ABSTRACT:

Diese Studie untersucht die Effektivität von Machine Learning-Algorithmen bei der automatischen Klassifizierung und Zusammenfassung von Textdokumenten.

### EINLEITUNG:

Die exponentiell wachsende Menge an digitalen Dokumenten erfordert innovative Ansätze zur automatischen Verarbeitung und Analyse.

### METHODOLOGIE:

- Datensammlung: 10.000 Dokumente aus verschiedenen Domänen
- Preprocessing: Tokenisierung, Stemming, Stopword-Removal
- Feature Extraction: TF-IDF, Word2Vec, BERT-Embeddings
- Klassifikationsalgorithmen: SVM, Random Forest, Neural Networks
- Evaluationsmetriken: Precision, Recall, F1-Score

### ERGEBNISSE:

- BERT-basierte Modelle zeigten die höchste Genauigkeit (94.2%)
- Random Forest erreichte 87.5% Genauigkeit bei geringerer Rechenzeit
- SVM zeigte robuste Performance bei kleinen Datensätzen (85.1%)

### DISKUSSION:

Die Ergebnisse zeigen, dass moderne Transformer-Architekturen wie BERT überlegen sind, jedoch erhebliche Rechenressourcen benötigen.

### LIMITATIONEN:

- Begrenzte Sprachunterstützung (nur Deutsch und Englisch)
- Domänenspezifische Bias in den Trainingsdaten
- Hoher Rechenbedarf für BERT-Modelle

### ZUKÜNFTIGE FORSCHUNG:

- Entwicklung effizienterer Modellarchitekturen
- Untersuchung von Few-Shot Learning Ansätzen
- Integration von multimodalen Daten (Text + Bilder)

### SCHLUSSFOLGERUNG:

KI-gestützte Dokumentenanalyse zeigt großes Potenzial für praktische Anwendungen, erfordert jedoch weitere Forschung zur Optimierung.

### REFERENZEN:

- [1] Smith, J. et al. (2023). "Advanced NLP Techniques"
- [2] Mueller, A. (2022). "Document Classification Methods"
- [3] Wang, L. (2024). "BERT Applications in Industry"