Goethe-Universität Frankfurt am Main
Fachbereich Wirtschaftswissenschaften

Professur für Betriebswirtschaftslehre,
insbesondere e-Finance

Prof. Dr. Peter Gomber

Theodor-W.-Adorno-Platz 4
RuW, Postfach 69
D-60629 Frankfurt am Main

Telefon +49-69-798-34683
Telefax +49-69-798-35007
E-Mail gomber@wiwi.uni-frankfurt.de

http://www.efinance.wiwi.uni-frankfurt.de

## Textual Content Generation with Recurrent Neural Networks

Advances in the realm of machine learning and specifically recurrent neural networks (RNNs) such as Long Short Term Memory (LSTM) allow for the algorithmic generation of textual content. The goal of this master's thesis is to implement a system based on RNNs to generate textual content that is (almost) indiscernible from content created by humans. To iteratively develop the system and to proof its applicability, the student is expected to run multiple experiments. First, the student will apply widely used statistical machine learning classification algorithms such as Logistic Regression, Support Vector Machines, Decision Tree, Random and Extreme Gradient Boosting on a balanced sample of texts generated by humans and texts generated programmatically by RNNs. The cross-validated classification performance is expected to be benchmarked against a simple Markov chain based pseudo random text generator. Second, the student is expected to run an experiment in which human evaluators will be asked to differentiate between texts generated by humans and texts generated by the RNN.

Because of the technical nature of this Master's thesis, previous (substantial) programming experience in a programming language (e.g. R, Go, Python) and a strong interest in statistical machine learning is a prerequisite.

### Literature:


- **Breiman, L. (2001):** "Random forests" in: Machine Learning, Vol. 45, No. 1, pp. 5–32
- **Chen, T. and C. Guestrin. (2016):** "XGBoost: A Scalable Tree Boosting System" in: ACM Press., pp. 785–794
- **Colah's blog (2015):** "Understanding LSTM Networks", http://colah.github.io/posts/2015-08-Understanding-LSTMs/ (Accessed on 12.10.2017)
- **Cortes, C. and V. Vapnik. (1995):** "Support-vector networks" in: Machine Learning, Vol. 20, No. 3, 273–297
- **Cox, D. (1958):** "The regression analysis of binary sequences (with discussion)" in: J Roy Stat Soc B, Vol. 20, pp. 215–242
- **Hochreiter, S. and Schmidhuber, J. (1997):** "Long short-term memory" in: Neural computation, Vol. 9, No. 8, pp.1735-1780
- **Kohavi, R. (1995):** "A study of cross-validation and bootstrap for accuracy estimation and model selection" in: Ijcai, Vol. 14, No. 2, pp. 1137-1145.
- **Quinlan, J. R. (1986):** "Induction of decision trees" in: Machine Learning, Vol. 1, No. 1, pp. 81–106.
- **Walker, S. H. and D. B. Duncan. (1967):** "Estimation of the probability of an event as a function of several independent variables" Biometrika, Vol. 54, No. 1–2, pp. 167–179
- **Williams, R.J. and Zipser, D. (1989):** "A learning algorithm for continually running fully recurrent neural networks" in: Neural computation, 1(2), pp.270-280.


### Supervisor: Christian Janze