

## SUMMARY

I am a Data Scientist by educational background. 2 years of experience in the telecommunications industry performing statistical analysis, predictive modeling, text analysis and deep learning. Have hands-on experience with managing stakeholders, storytelling recommendations, coaching a team on Python data analysis, productionizing models, and A/B testing.

## EXPERIENCE

- Customer Experience Analyst** VodafoneZiggo | Utrecht | Dec 2018 - Present  
Deriving actionable insights and creating predictive models for customer experience improvement.  
Key achievements:
  - Built and deployed a SparkML model to predict calls from traffic on the company website and proactively offer a chat instead to enable digital channel shift. Saves between ~750-2000 calls per week based on chat engagement rate. Improvements to the model are A/B tested.
  - Applied statistical analysis on signal data to understand effect of troubleshooting guides on modem signal and suggested recommendations.
  - Performed Regression analysis on NPS surveys to inform journey managers on where to focus improvement efforts.
  - Applied text analysis on agent logs to determine more specific call reasons, trending topics over time.
  - Pretrained Google's BERT on B2C customer service chat data and applied towards company-domain specific keyphrase extraction.
  - Coached my team on Python based analysis, regression modeling, text analysis.keywords: Python, Spark, SQL, ML, NLP, Statistics, LightGBM, Jupyter Notebooks, Qlik
- Data Science Intern** ING Bank | Amsterdam | Sep 2017 - Sep 2018
  - Graduation Thesis: Research and design of novel framework for predicting channel preference of customers to increase conversions and return-on-investment on marketing campaigns.
    - Translated business requirements of database marketers into model design.
    - Formatted, cleaned, combined and filtered billions of rows of customer channel interactions data.
    - Engineered >50 features to quantify customer behavior across online and offline channels.
    - Implemented a production-ready three-stage ML framework to increase customer responses.
    - Promoted the product to business stakeholders.
  - Internship project: Design of new banking leads generation model to improve leads quality.
    - Discovered and fixed critical issues in feature generation process in existing deployment.
    - Leveraged feedback loop data and created a production-ready model to improve precision of predicted leads.
    - Created a model workflow simulation to give proof of continuous improvement to stakeholders.
    - Forked, improved and deployed a python library to distribute scikit grid searches over spark nodes minimizing training time.keywords: Python, Spark, Bash, Hadoop / Hive, Netezza SQL, Git
- Research associate** Blue Mango | Eindhoven | Sep 2015 - Sep 2016  
Researched algorithms to improve predictions of Ad Click-through rates (CTR) central to the online marketing business of Blue Mango.
  - Solved the main challenge of class imbalance due to low CTR in advertising domain.
  - Benchmarked multiple algorithms with and without applying SMOTE upsampling techniques.
  - Achieved 59% decrease in average log-loss from 0.68 to 0.28 on real-world datasets with SMOTE + random under-sampling, thereby confirming research hypothesis.keywords: Python, R, MySQL, Git

## EDUCATION

- MSc. Computer Science Engineering** Eindhoven University of Technology | 2016 - 2018 | GPA: 8.1 / 10
  - Specialization in Data science and Software Engineering.
  - Excelled in machine learning and data science coursework.
- BSc. Computer Science Engineering** Eindhoven University of Technology | 2013 - 2016 | GPA: 7.9 / 10
  - Bachelor Thesis: developed a web application to connect football coaches and players for organizing club activities. Achieved a 9.5 (of 10) with recognition for problem solving skills and team work under Agile setting.
  - Specialization in Software Engineering and electives in AI and statistics.
  - Well-versed in Agile scrum practice.

## PROJECTS

- **topic-analysis** Aug 2020 - Present  
Library to gain fast insights from text data with topic analysis and visualize trending topics.
  - Extract noun phrases with POS tags, vectorize with BNgrams, cluster phrases based on mutual information into topics and assign documents to topics.
  - Visualize trends of generated topics over time.
  - Based on the paper: Mining newsworthy topics from social media
- **spacycaKE** Jul 2020 - Jul 2020  
Simple keyphrase extraction extensions and pipeline components for spaCy.
  - Based on the paper: Simple Unsupervised Keyphrase Extraction using Sentence Embedding
  - Link to implementation: <https://github.com/surajiyer/spacycaKE>
- **spacybert** Jul 2020 - Jul 2020  
BERT inference for spaCy with custom extension attributes
  - Provides a simple spacy compatible interface for extracting sentence embeddings from BERT model(s) for downstream application. Uses the Hugging Face Transformers library as backend for loading Bert models.
  - Link to implementation: <https://github.com/surajiyer/spacybert>
- **python-data-utils** Jan 2019 - Present  
Utility functions and classes for many common data science libraries.
  - A collection of standard implementations of common algorithms and helper functions used in Data science workflows.
  - Link to implementation: <https://github.com/surajiyer/python-data-utils>
- **adidas Hackathon 2018** Jun 2018 - Jun 2018  
Personalizing & improving next-generation shopping experience by integrating machine learning.
  - Developed generative adversarial networks (GANs) to create personalized products by "transferring the style" of your favorite patterns / textures / images.
  - Developed an independent hybrid mobile application with IONIC framework which communicated with the prediction backend, bringing our team to the finals.
  - Link to implementation: <https://github.com/surajiyer/Adidas-Hackathon-2018>keywords: Python, Tensorflow, IONIC, MEAN stack, Git
- **Recommender Systems** Mar 2018 - May 2018  
Tutorials on applying neural networks to information retrieval challenges.
  - Using CBoW and Skipgram models to generate word embeddings for analogy detection and sentence reconstruction.
  - Generate image neural codes using AutoEncoders and apply them to image retrieval.
  - Using Siamese networks and RNNs to generate image and document embeddings respectively and apply them to one-shot learning tasks.
  - Exercise on using Triplet network for unsupervised image-caption retrieval problem.
  - Link to IPython notebooks: <https://github.com/surajiyer/Recommender-Systems>keywords: Python, Keras, Matplotlib
- **Transfer learning with TrAdaBoost** Jan 2018 - Jan 2018  
Developed a python scikit-learn implementation of the TrAdaBoost algorithm from ICML '07 "Boosting for Transfer Learning".
  - Exploring transfer learning as a solution to sample selection bias correction in training models on financial datasets.
  - Link to implementation: <https://github.com/surajiyer/Transfer-learning-with-TrAdaBoost>
  - Link to the paper: <http://www.machinelearning.org/proceedings/icml2007/papers/72.pdf>keywords: Python, Sklearn
- **Brexit Rumor Analysis** Sep 2016 - Nov 2016  
Rumor detection and verification system for the Brexit referendum topic.
  - Setting up data collection and text preprocessing for >1 million brexit tweets.
  - Engineered 30+ tweet-level, user-level and rumor-level features per tweet.
  - Helped with hierarchical clustering of tweets into predefined rumor topics using topic filters.
  - Designed two models: (1) for verification of rumor, (2) for falsification of rumor. Combine them both using natural cut-off values to control precision and recall.
  - Link to implementation: <https://github.com/mthaak/2imw15-project>keywords: Python, Sklearn, ElasticSearch

## SKILLS

### PROGRAMMING

Python  
SQL  
NoSQL  
R  
Java  
HTML / CSS / JS

### TOOLS & LIBRARIES

Bash      Matplotlib  
Git        Statsmodel  
Spark     Keras  
Pandas    Node.js  
Numpy  
Pytorch  
Scikit-learn

### CORE SKILLS

Machine Learning  
Statistical Analysis  
Text Analysis  
Software Engineering  
Research

### SOFT SKILLS

Storytelling  
Project Leadership  
Managing stakeholders

### LANGUAGE

English      Professional  
Dutch        Elementary