Diego Bettega

Data Scientist

Summary

+39 324 7728599

diego.bettega.90@gmail.com

104 Consolatico Superiore,
Busseto (PR), 43011, Italy

LinkedIn Profile

GitHub Repository

Machine Learning & AI

Python Packages:

- Numpy, Pandas, Matplotlib, Seaborn
- Scikit Learn, Scipy, TensorFlow, Keras

Supervised Learning:

- Regularised Linear & Logistic models
- Naïve Bayes, KNN, Decision Tree, Kernel SVM
- Ensemble techniques: Bagging, Random Forest, Adaboost, XGBoost, Light GBM, Voting Classifier

Unsupervised Learning:

- K-means, Hierarchical Clustering

Dimensionality Reduction:

- PCA, Kernel PCA, LDA, SOMs, t-SNE, UMAP

Recommendation Systems:

- Market Basket Analysis, Collaborative Filtering
- SVD, RBM

Deep Learning:

- ANN, CNN, RNN-LSTM, GANs, AutoEncoders

Reinforcement Learning (limited experience):

- UCB, Thompson Sampling
- Markov Decision Process

Skills

Python	Good
Cloud & Big Data	Medium
MySQL	Good
alteryx - certifications	Advanced
Tableau	Good
DevOps	Medium

Languages

Italian	Native
English	Good

Work Experience

Mar 2018 - Present

Snr Business Analyst / Data Scientist, EMEA

Kinesso (IPG) – London (UK) & Milan (IT)

- Pioneered and build a global model for automatic marketing campaign optimizations, using machine-learning models (Python) and deployed in a React JS Web App. Campaigns costs were reduced by 30% on average.
- Automated a reporting process by collecting data from multiple Google marketing sources (ADH) and Amazon databases (AMC) using APIs and SQL queries, analysing the data in Python and visualizing the results in Tableau. Implemented Bidirectional LSTM model for path-to-conversion analysis and ML models for optimal frequency cross-campaigns and insights.
- Successfully designed and implemented an audience insights Tableau dashboard, empowered by first party, third party and programmatic data encoded by using R and alteryx and queried using MySQL. This new product is currently used and sold by EMEA senior leadership team.
- Promoted three times in the first three years.

Jul 2016 - Feb 2018

Data Analyst

Adloox – London (UK)

- Increased revenue, review discrepancy and detect different types of inefficiency based on Brand Safety, Viewability & Fraud Detection criteria.
- Automated time-consuming manual reporting processes that could easily lead to errors (VBA Excel Macro).

Education

Jun 2019 - Jul 2020

The University of Texas at Austin *#2 in Analytics, #4 in AI, #7 in Machine Learning*

PGP in AI and ML *(Grade: Excellent)*

Advanced statistic; supervised - unsupervised - reinforcement learning; featurization, model selection and tuning; recommendations systems; deep learning (ANN, CNN, RNN - LSTM); NLP; GANs; model deployment.

Winner of a Hackathon with 120 data scientists participating in the competition. The objective was maximizing the accuracy in a supervised classification problem.
#4 place in "The Grand Hackathon" with over 1000 data scientists participating in the competition. Won a prize: course in Mastering Big Data Analytics.

Jun 2020 - Feb 2021

PGP in Cloud Computing

Cloud Foundations; Specialization in AWS, Microsoft Azure and Google Cloud; Containers; Microservices; Big Data Management and Analytics on Cloud; Cloud Security & Migration; Private Cloud; Enterprise Cloud Solutions; Cloud-Native DevOps; On Prem DevOps.

Dec 2020 - Present

Course in Mastering Big Data Analytics

Apache Hadoop; Map reduce; HDFS; YARN; Hive; Pyspark; Spark SQL; Spark MLIB; Spark streaming; Kafka.

Dec 2014

Mechanical Engineering

Bachelor's Degree – University of Parma, IT

Projects

- Financial model: ensemble techniques and stateful LSTM on FOREX market prediction.
- NLP: GloVe embedding and multi-layered bidirectional LSTM on sarcasm detection.
- Face Detection: CNN (UNET architecture) to locate the position of a face in an image.
- Face Recognition: pre-trained VGG face, Triplet Loss, PCA and SVM to recognize faces.
- Image Classification: ANN in Numpy & Keras to identify images containing numbers.
- Recommendation System: recommendations for products on an e-commerce website.
- Customer Retention: stacking multiple ML model to identify churning customers.
- Model Explainability: partial dependence plot and SHAP plot to generate insights.
- Team Communication: Team Communication Solution.
- Managed Service: Automated Process using Managed Services on a Public Cloud.
- Web App: Web App to ECS.
- Master-less arch concepts: Install multi-node Cassandra cluster and induce failure.
- Jenkins Server: Remote manage EC2 instances and Build a Jenkins server.