Marc ALPHONSUS

Machine Learning Engineer, 8+ years of experience

Senior data scientist with a portfolio ranging from scoring to fraud detection, with experience in image recognition and language modeling in agile and data driven environments. I'm looking for either a Tech Lead role in a team of Data Scientists or a senior Data Scientist position. About me : I am a product-oriented data nerd with a teamplayer mindset infused to the devops philosophy, my favorite perks in the job are the needed creativity and abstraction during model design and industrialization.
If any, my expertise revolves aroung Deep Learning in particular NLP and reinforcement learning. Also I love code, kaggle, coffee and comics (order may vary).
#NLP #activelearning #image #eli5 #rl #federatedlearning

Niort, France
32 ans, french
Car and motorcycle driving licenses.

alphonsus.marc@gmail.com
linkedin / github
+33 6 62 62 62 48

📉 Experience 📈

CGI 🧢 Lead Data Science 🗓️ 2018 - 2021

Manage the AI lab agency while offering a window to our projects and skills Machine Learning wise. Define and build the data strategy of my clients.

IMA :

- Assist+: PoC for a AI phone assistant - combined ASR/STT and various NLP models (NER, Classification, Semantic Similarity) to better the assistance process when facing car failures and accidents;
- Process : Speech transcription, vehicle identification, geolocalization, failure understanding from client's description, services (motor mechanic, taxi, hotel…) & SMS recap > `pytorch, Go, docker`

MACIF :

- Macif NLP: created models for all Macif's textual content (routing/classification, named entity recognition, sentiment analysis)- bumped up the performance from 85% over 3 classes to **92% over 56 classes** > `Tensorflow, Huggingface transformers, spacy, docker`
- Hub IA: built the group's AI services platform, monitored and secure, as a swagger API catalog callable by every employee > `WSGI, Flask, FastAPI, Docker, Kubernetes`
- AutoML Workflow : assembled an automatic AI retrain pipeline (similar to `MLFlow`) - Models are being automaticly retrained with new incoming data fed thanks to an users feedback loop (active learning, multitask-learning) > `python, gitlab CI, Docker`
- RAD/LAD: constructed a service for analyzing documents (classification & OCR) before storage **95% over 80 classes** > `pytorch, Docker`
- MeetUps: defined the Data strategy at the departement level & AI evangelization for the whole company `OKR, Gslides, Jupyter, Streamlit`

- Chatbot EPI: redesinged the UX of the online member space (4.3 M users) - the deployment of AI models allowed to meet key objectives while modernizing the overall architecture > `Tensorflow, spacy, rasa`

CGI - AI Lab :

- Project management & mentoring
- SOTA and technological monitoring
- Conferences, ELI5

MAIF 🧢 Data scientist 🗓️ 2014 - 2018

Bring to the various Departments funding the corresponding data projects the required expertise in statistical learning while conceiving the future tools to serve the MAIF members.

- Mail: Automatic routing of e-mails & summarization, open sourced as Mélusine ; design of neural architectures and extractives summarization algorithms > `tensorfow, pySpark, scikit-learn, docker`
- MAIF & Go: Claim pricing based on driving behaviours obtained by GPS trackers (IoT beacon & mobile app) > `pySpark, R, caret`
- Valeur Sociétaire: * Modeling of expected rentability for customers on a large temporal horizon (LifeTime Value)* > `pySpark, Hadoop Map/Reduce`
- Téléphonie: Predict the phonecall load to correctly size call centers HR wise **5% error, publication proposal** > `R, caret`
- risk: Identity under-insured risks > `scikit-learn`

Poitiers University 🧢 Machine Learning TA 🗓️ 2017 - 2019

Introduce grad students to machine learning

- Syllabus: decision theory, regression, SVM, CART, random forests, MLP and convolutionnal neural networks.

MAPA 🧢 Statistician 🗓️ 2013 - 2014

Analyze and explain pricing drifts, price elasticity and financial performance.

🛠️ Hard skills & extra curriculum ⚙️

Chosen projects	**TextNets/text**: My own models and tests regarding SOTA NLP, paper replication, open source model training, semantic similarity, automatic summarization, transduction, search… **DeepCount**: Crowd counting with density mapping on highly populated images (think football game)
Hackathons	**Hackathon IMA**: top 1 **AfLPC x LineUp7**: top 2

Languages: `Python, R, Go, SAS`
Spoken Languages: `English, Français, Deutsch, 日本語 (notions)`
Frameworks: `pytorch, tensorflow, scikit-learn, Apache Spark, Hadoop, opengym`

Cloud: `AWS, GCP`
Basics: `UNIX, algorithms, git, gitlab-ci, Docker, redis, MongoDB, SQL`

Education & Training

2020	**Spinning up Reinforcement Learning**; OpenAI **6.S094**; MIT
2018	**CS224n, CS229, CS231n**; Stanford University
2017	**Deep Learning Specialization**; Deeplearning.ai x Coursera
2014	**Machine Learning**; Stanford University x Coursera **Deep Learning**; Google x Udacity
2013	**MSc, Statistics**; Université de Poitiers
2010	**Bachelor, Mathematics and Computer Science**; Université Paris V