James Morris, PhD

Data Scientist

me@james-morris.com
347-384-8967
james-morris.com
linkedin.com/in/jamespmorris
www.github.com/jpmorris
stackoverflow.com/users/795276

Exceptional Problem Solver • Highly technical Practitioner • Superb Hacker • Strong Leader

Detail-oriented, broad-perspective researcher with more than a decade of experience with various methods of analysis and modern technologies. Expert at developing solutions, communicating results and delivering value. Broad experience with AI (with a strong emphasis on classical and deep machine learning), data mining, classical statistics, and physical and mathematical modeling. Strong programming skills in algorithms, data structures, software development, and high-performance computing. Hands-on experience with the full stack of data science technologies. Excellent communicator, skilled in project management, and experienced team lead. Seeking lead position as a practitioner solving very hard problems.

Degrees in physics, math, philosophy, computer science • Data Science, Development, Architecture, Engineering, Project Management • Python, R, Spark, Java, C#, OOP • Classical & Deep Learning, AI, Advanced Analytics

Recent Experience

INFOSYS — Silver Spring, MD
Senior Data Scientist – Practice Lead — 03/2015—Present

For a multinational oil & gas supermajor:

- Engineered machine learning model using high-frequency, real-time data with scikit-learn and other deep LSTM models using Keras/Tensorflow, enabling a savings of at least USD 10 million/year by preventing drilling failures
- Helped lead team to architect and build real-time high-frequency low-latency data pipeline within Azure to enable USD 10 Million/year
- Hands-on experience with big data infrastructure technologies: C#/.NET, NIFI, Kafka, Databricks, Spark, Kubernetes, Azure App Service, Cosmos, MongoDB among others
- Created deployment pipeline using DevOps CI/CD to deploy to Azure Kubernetes Service
- Hands-on experience with big data machine learning platforms including Azure ML, MLflow
- Developed a GTK3 application to obtain consistent set of data labels from SME to apply to the dataset at scale
- Detected anomalies in dirty hydraulic fracturing data by general engineering and mining the frequency space using Fourier decomposition, dimensionality reduction methods including PCA, and clustering (k-means, DBSCAN)
- Developed code to optimize the chemical inputs of processing plants using sequential least squares programming

For a leading financial services company:

- Developed model to predict smoking propensity as part of a rapid underwriting system
- Created sales-lead lift model using R (Caret) for multi-model comparison with C5.0 as the final model
- Built client prediction model using 3rd-party data broker to target individuals on Facebook
- Created revamped sales-propensity model using SAS EM
- Created turnover, retirement, and migration logistic regression model and library in R to mitigate up to USD 6 million lost annually in employee replacement
- Developed NLP models for text mining for sentiment analysis

For a multinational auto manufacturer:

- Created model in R to predict driver destinations from telemetric data

GEORGIA INSTITUTE OF TECHNOLOGY — Silver Spring, MD
Student — 09/2015—08/2020

- Developed AI algorithms to solve problems adversarial searches (minimax, alpha-beta pruning), graph searches (BFS, DFS, A*), Bayesian networks, Gaussian mixture models applied to image compression, among others
- Programmed AI methods to solve Raven Intelligence Matrices
- Created reinforcement solutions to Grid World via value and policy search methods
- Mined Wikipedia data for a knowledge ontology using Spark on Databricks (GraphFrame) and deployed a CI/CD-supported web app on AWS (CodeDeploy, CodePipeline, ACS, and Fargate)

OHIO STATE UNIVERSITY, DEPARTMENT OF PHYSICS — Columbus, OH
Research/Teaching Assistant — 09/2002—07/2009

- Using C++ and ROOT, performed a statistically blind analysis using Monte Carlo based simulations developed from first principles. To test the Standard Model of particle physics, the resulting model processed data from the world's first petabyte database in search of rare decay modes of fundamental particles
- Applied machine learning techniques (FDA, Fisher discriminates, and decision trees) in data mining to separate signal from noise (ROOT TMVA)
- Research fellow at INFN (Padua, Italy) and KEK (Tsukuba, Japan), providing research assistance for infrastructure and simulation support
- As sub-detector quality control expert and software release coordinator, used Perl to automate tasks
- Extensive experience communicating complex ideas to experts and neophytes: teaching assistant winning multiple awards, presented at multiple conferences and regularly to experts

Education

PhD – Physics
Ohio State University — 08/2002—07/2009

MS – Computer Science
Georgia Institute of Technology — 08/2015—08/2020

BS – Physics/Mathematics
University of Arizona — 09/1997—04/2002

BA – Philosophy
University of Arizona — 09/1997—04/2002

Medicine
Stony Brook University — 08/2009—11/2013

Skills

Soft

Communication	❶❷❸❹❺
Problem Solving	❶❷❸❹❺
Leadership	❶❷❸❹⑤

Neighbor

Engineering	❶❷❸❹❺
App Development	❶❷❸❹⑤
Project Management	❶❷❸❹⑤
Data Architecture	❶❷❸❹⑤

Analytic

Classical Learning	❶❷❸❹⑤
Advanced Analytics	❶❷❸❹⑤
Mathematics	❶❷❸❹⑤
Statistics	❶❷❸④⑤
Simulations	❶❷❸④⑤
Deep Learning	❶❷❸④⑤
Broader AI	❶❷❸④⑤
NLP	❶❷❸④⑤
Reinforcement Learning	❶❷③④⑤

Languages

Python	❶❷❸❹⑤
R	❶❷❸❹⑤
SQL	❶❷❸❹⑤
Spark	❶❷❸④⑤
C++/C#	❶❷❸④⑤
Java	❶❷❸④⑤
JavaScript	❶❷❸④⑤

Computer Science

Algorithms	❶❷❸❹⑤
High-Perf Computing	❶❷❸④⑤

Big Data

Azure	❶❷❸❹⑤
AWS	❶❷❸④⑤

Hacking

Solution Engineering	❶❷❸❹⑤
Dirty Data	❶❷❸❹⑤
Getting Data	❶❷❸❹⑤

Domain

Energy	❶❷❸❹⑤
Financial	❶❷❸❹⑤