Farrukh Nauman

farrukhnauman@gmail.com | (+46) 0702984959 | LinkedIn: fnauman | Github: fnauman
Honnörsgatan 1 LGH 1405 Linköping 58752 (Sweden)

Summary: As a Machine Learning researcher holding a PhD in Physics, I specialize in the integration of traditional scientific methodologies with advanced machine learning algorithms. At present, I am deeply involved in two large-scale sustainable fashion initiatives. The first major contribution of these projects is the creation of a unique, annotated dataset for damaged used clothes, setting a new industry standard. The second entails the development of specialized AI models, leveraging cutting edge vision algorithsm to identify damage and multi-modal generative AI models to describe the clothes for customer to customer websites. I have experience working in, and leading, diverse teams of various sizes and have mentored junior developers and students. I am seeking senior research and data science roles.

Skills

- **Programming languages**: Python (6+ years), SQL (2 years), C (7 years), Fortran (3 years), IDL (3 years).
- **Languages**: English (fluent), Swedish (beginner), Urdu (native).
- **Machine learning**:
  - PYTORCH: Primary framework for most of my deep learning based work, also wrote a fully differentiable CFD solver in PYTORCH.
  - SCIKIT-LEARN, PANDAS: Classical ML algorithms.
  - SKTIME, DARTS: Time series forecasting and classification algorithms.
- **Algorithms, DevOps**: Mulit-modal (Vision-Language) Models and Generative AI, Azure data engineering and model deployment.

Experience

Research Institutes of Sweden (RISE) — Linköping, Sweden
AI Researcher — *Jul 2021 -*

- **AI for resource-efficient fashion**:
  - Two large projects (Vinnova, cisutac.eu) with the goal of making the fashion industry more sustainable. These projects were only 1 of 5 projects to be presented at EU event on sustainability and AI, May 2023 and Vinnova Innovation week (Sep. 2022).
    - Dataset: We released version 1 in Sep. 2023.
    - Algorithms:
      (1) Self-supervised vision (MAE) and multi-modal models (CLIP) to learn image representations for image similarity analysis (Master's thesis).
      (2) Generative multi-modal models OpenFlamingov2, Grounded SAM and ControlNet to (i) detect and segment damage like holes and stains in clothes, (ii) image editing: remove damage and estimate new price.
    - Goals:
      (i) Sorting: Decide whether a clothing item is more suited for reuse/resale, recycling, repair, export, waste.
      (ii) Attribute prediction: price, trend, damage and other attributes.
      (iii) Create a dataset of damaged clothes with annotations.

    Overall, my role has been to develop everything from a simple (streamlit and flask) annotation app that can interface with two cameras, to training machine learning models to generating and presenting reports to stakeholders.
- **Other**:
  - Project - Low Energy IoT: AI algorithms and models suitable for analyzing sensor data both offline and online.
    - Report on energy efficient AI models that are suitable for edge devices.
    - Analysis of sensor data from human runners.
    - Report on the state-of-the-art time series classification algorithms.
  - Mentorship: I have mentored several Master's thesis in AI.
    - Computer Vision: Simon Hermansson - Self-supervised learning, Theodor Emanuelsson - Object detection (Jun. 2023).
    - Time Series Forecasting: Ongoing thesis work on forecasting fashion trends using the multi-modal Visuelle 2.0 dataset.

2MNordic IT Consulting AB — Gothenburg, Sweden
Data Scientist — *Dec 2019 - Jun 2021*

- **Helsingborg secondary schools:** :
  - Project I: Early warning system for students and funding prediction for schools:
    - Identified core subjects and number of fails in 6th grade as the primary indicators of performance in 9th grade.
    - Identified most important features schools that require funding to improve school-wide performance.
    - Clustered similar students and schools together to gain insights into important features.
  - Project II: Analyze student reviews and new digital Mathematics test
    - Found inconsistencies between the existing heuristic grading with the new Mathematics test.
    - Discovered that higher student reviews about the school and teachers correlate negatively with student grades.
  - Reports: Prepared Power BI report for the stakeholder.
  - Cloud Deployment (current): Using Azure Functions and Data Factory.
  - Tools: Azure DevOps, Power BI, SCIKIT-LEARN, PANDAS, Azure Functions and Data Factory.

- **Cloud certifications**: Microsoft Azure Data Engineer Certificate.

Chalmers University of Technology Gothenburg, Sweden
Origins Fellow (Machine Learning, Astrophysical Fluid Modeling) *Sep 2018 - Nov 2019*

- **Machine Learning models of high dimensional fluid and plasma simulations**:
  - Studied magnetic field growth in turbulent plasmas using time series data.
  - Surrogate modeling of complex ODEs/PDEs using recently proposed algorithms like SINDy.
- **Planet formation**:
  - Set up large scale plasma simulations to understand early stages of massive star formation.
  - Tools: C for high performance computing; Python for post-processing.

Niels Bohr Institute Copenhagen, Denmark
Postdoctoral Fellow (Astrophysical Fluid Modeling) *Sep 2015 - Sep 2018*

- **TransitionToTurbulence**: A systematic study of the importance of aspect ratio in the transition to turbulence.
- **LocalDisks**: Used fluid simulations (SNOOPY) to model astrophysical turbulence.

University of Rochester New York, USA
Research Assistant/PhD Student *Sep 2009 - Oct 2015*

- **Accretion disk simulations**: Designed and tested Computational Fluid Dynamics simulations to study accretion flows.
- **Data analysis**: Analyzed data in the excess of 10 TeraBytes using IDL, MATLAB for publication of 4 papers.

EDUCATION

University of Rochester Rochester, New York (USA)
PhD in Physics and Astronomy *Oct 2015*

Thesis: Turbulence in Rotating and Non-Rotating Magnetohydrodynamic Shear Flows.

Quaid-i-Azam University Islamabad, Pakistan
M. Phil. Physics *June 2009*

Thesis: Modified gravity as an explanation for cosmic acceleration.

AWARDS AND & ACHIEVEMENTS

- Horton fellowship from Laboratory for Laser Energetics. 2010-2015
- Susumu Okubo Prize for the highest performance on the graduate physics written comprehensive exam and excellence in coursework. 2011

SELECTED PUBLICATIONS

- **Reducing the complexity of chemical networks via interpretable autoencoders**: Tommaso Grassi, **Farrukh Nauman**, et al., 2020. A&A . (arxiv.org/abs/2104.09516):
  - Reduced a big astrophysically relevant chemical network ODE system to a computationally tractable ODE system.
  - Used regularized autoencoders with ODE solvers in TENSORFLOW.
  - Designed the algorithm with custom layers and loss functions.
- **Machine Learning models of magnetic field generation**: **Farrukh Nauman**, Joonas Nättilä. 2019. A&A 629, A89. (arxiv.org/abs/1905.08193) (Blog: fnauman.github.io, Code: github.com/fnauman/ML_alpha2):
  - Used decision tree and Bayesian algorithms to analyze time series data from magnetic field generation simulations.
  - Using the right features, linear regression outperformed more complex algorithms.

REFEREEING SERVICE FOR JOURNALS

Machine Learning Science and Technology, Astrophysical Journal, Monthly Notices of Royal Astronomical Society, Journal of Cosmology and Astrophysics, European Physical Journal Plus