**Ehsan Hesamifard**

**Tallahasse, FL, 32311 | Email: ehf.ehsan@gmail.com**

**PROFILE**

Highly qualified and tenacious machine learning expert with broad-based skills and extensive experience in machine learning and data science, artificial intelligence, data privacy, computer programming, statistical modeling and database management. Combines broad machine learning background with solid data science acumen to implement solutions that solve business problems. My Ph.D. dissertation focuses on Machine Learning & Data Privacy and I possess outstanding machine learning expertise and data science proficiency to assume the position of a **Data Scientist.**

**COMPETENCIES**

* Advanced written and verbal communication skills.
* Strong expertise in data privacy, data analysis, machine learning, statistical modeling, and AI Explainability.
* Extensive use of Python and proficient in Python, C, and C++.
* Adept at working with Keras, WEKA, scikit-learn, MATLAB.
* Adept at working with Spark, PySpark, and slurm.
* Adept at working with databricks.
* Adept at working with Tableau.
* Experience working with Tensorflow, PyTorch, XGBoost and R.
* Self-sufficient in applications of homomorphic encryption in data privacy.
* Extensive use of classic machine learning algorithms, such as logistic regression, SVM, Decision Trees, Naïve Bayes, Random Forest, etc.
* Great command of statistical analysis techniques and building predictive models such as K-Means clustering, choice modeling, and association rule.
* Strong familiarity with SQL to manipulate data, and in-depth knowledge of Hadoop, Spark, and Scala.
* Collected data from various RDBMS and converted raw data to processed data or tidy data.
* Ability to present complex data and analytics to the non-analytical audience.
* Provide guidance and mentor junior analysts (REU and Tams programs).

**Skills and EXPERTISE**

* Natural Language Processing
* Optimization Algorithms
* Computer Vision
* Artificial Intelligence
* AI Explainability
* Statistical Analysis
* Data Visualizations
* Statistical Modeling
* Machine Learning
* Data Manipulation
* Neural Networks
* Data Privacy
* Data Science
* Data Analysis
* Cryptography

**WORK EXPERIENCE**

**Complaint Tracer. Jan 2024 – Dec 2024.**

* Utilized my expertise in NLP and GenAI technology to build a complaint identification solution.
* Developing service the solution as a service.

**Project description**:

* Model the problem within GenAI limitations.
* Prompt Engineering.
* Using Llama to generate the output.

**Environment: python, Cortex, Snowflake.**

**Chatbot. Jan 2023 – Dec 2023.**

* Utilized my expertise in NLP and GenAI technology to build a RAG base chatbot.
* Developing UI for pro-type.

**Project description**:

* Process text data in scale.
* Implementing solution based on GenAI models such as Llama, Falcon, Mixtral.
* Implementing pipeline from loading data, processing, text similarity and generate output based on GenAI model.
* Run model as an API endpoint.

**Environment: python, Text-Generation-Inference service, Flask, streamlit.**

**Text Similarity Model. Sept 200 – Sept 2022.**

* Utilized my expertise in NLP and classification algorithms to build text similarity models.
* Building recommendation system for find best match for input texts.

**Project description**:

* Extract data from Hadoop.
* Process steps such as text processing and cleaning.
* Building models based on different techniques such as string matching, w2v and text similarity algorithms.
* Explain results and deliver model.

**Environment**: python, pyspark, SQL, Jupyter Hub, NLTK, Spacy, tensorflow.

**Model Deployment Prototype. Sept 200 – Sept 2022.**

* Utilized my expertise for building model deployment prototype.
* Developing end-to-end pipeline for deploying models.

**Project description**:

* Airflow for task orchestration.
* Mlflow for model lifecycle managent.
* Slurm for running task in a distributed system.

**Environment**: python, pyspark, SQL, Jupyter Hub, airflow, mlflow, slurm.

**Customer Classification Model. Sept 2021 – Sept 2022.**

* Utilized my expertise in deep learning and classification algorithms to build text similarity models.
* Building classifier for classifying risk associated to clients.

**Project description**:

* Extract data from Hadoop.
* Process steps such as data processing, EDA and data visualization.
* Extract features from data, using different feature analysis for feature engineering.
* Building models based on neural network and XGBoost.
* Using AI explainability tools such as SHAP for explaining results.

**Environment**: python, pyspark, SQL, Jupyter Hub, tensorflow, SHAP.

**Customer Portofolio. Sept 2020 – Sept 2022.**

* Utilized my expertise in clustering and classification algorithms to cluster customers.
* Building recommendation system based on the trained model.

**Project description**:

* Extract data from Hadoop.
* Process steps such as cleaning and EDA.
* Extract features from data.
* Using clustering/classification methods to build models.
* Explain results and deliver model.

**Environment**: python, sklearn, pyspark, SQL, Jupyter Hub, bokeh.

**Metadata Quality Assesment. Sept 2020 – Sept 2022.**

* Utilized my expertise in NLP, clustering, and classification algorithms to assess quality of metadata in databases.

**Project description**:

* Load data by using SQL query.
* Process steps such as cleaning extract data.
* Convert text to vector and extract features from text.
* Using clustering/classification methods to build models for assessing the quality of text.
* Run the model periodically to classify data.

**Environment**: python, tensorflow, sklearn, spark, SQL, Jupyter Hub, NLTK.

**Extract data from PDF. Sept 2020 – Sept 2022.**

* Read pdf documents in python and extract data.
* Export information from PDF as structured data.

**Project description**:

* Read PDF documents: using python to read PDF documents, process data and extract tables from the file.
* Process data in python and export as structured data.

**Environment**: Python.

**Image Processing - Overstock.com, Inc. Jun 2019 – Aug 2019**

* Utilized my expertise in Deep Learning for Image Quality Assessment, processing images and trained models with convolutional neural networks.
* Analyzed the manipulation methods for images like Gaussian Blur, Contrast, Quantization, Compression etc. and implementing them in Python.

Implemented final model over noisy dataset and improve the performance of the model.

**Project description**:

* Image Quality: extracting pattern noise in images like blurry or low contrast and simulate the same noise with image operations.
* The main challenge is adding noise to the images and simulate real noise with artificial methods, preprocess the data, and training the DL algorithm for quality assessment of images.
* Trained CNN models based on the new preprocessed images structure and improve the performance of the model by tuning hyperparameters related to the preprocessing methods and the DL algorithm.

**Environment**: Keras, Tensorflow, Scipy, Skimage, Python.

**Privacy-Preserving Deep Learning. Apr 2015 – Marh 2020**

* Utilized my expertise in ANN and trained models with neural networks and convolutional neural networks.
* Analyzed performance of the model and improve it.
* Implemented final model over encrypted data.

**Project description**:

* Machine learning over encrypted data: implementing different algorithms such as Neural Networks and Convolutional Neural Networks.
* The main challenge is adopting the ML algorithm within the limitation of HE schemes, HE libraries only support addition and multiplication (replacing ReLU and Sigmoid with polynomials, MaxPooling is not supported).
* Trained models based on the new structure and improve the performance of the model by tuning hyperparameters.
* Ran algorithms over encrypted data which includes learning from encrypted data and classifying encrypted instances.
* Trained different CNNs (VGG-16, VGG-19, AllConv and AlexNet) over different datasets such as MNIST, CIFAR-10 and CIFAR-100.

**Environment**: Keras, Homomorphic Encryption Library (HELib, SEAL, HEAAN, PALISAIDE), Python, C++.

**Data Mining Aug 2015 – Dec 2015**

* Prepared datasets, noise cleaning and maintained missing data.
* Implemented complete pip-line of data mining like feature extraction, running different algorithm (SVM, Decision Trees, Naïve Bayes, Random Forest, etc.), and tuning parameters for them.
* Decided the final algorithm, measured the performance of the final model and prepared the report.

**Project description:**

* Data analyzing of different datasets. Analyzed the impact of noise on the performance of data mining algorithms.
* Added different level of white noise to the dataset and measured the resistance of different models against noise level.

**Environment:** Python, WEKA, Keras, scikit-learn

**Data Mining Aug 2015 – Dec 2015**

* Extracted dataset’s key characteristic, such as class distribution, a variety of feature types, etc.
* Trained different ML algorithms over the dataset and evaluated the performance of models
* Analyzed the model performance to find the relation between datasets characteristics and performance of the model, such as noise in the features and labels.
* Improved the performance of models by changing feature types, cleaning up feature values, feature engineering or selection, address any noise in the class labels, etc.
* Analyzed the relation between features (attributes) and performance of the model, such as class distribution, distinct values, unique values for a nominal-valued feature, the interaction between pairs of variables (calculate correlation coefficient) for features and classes.
* Compared the effectiveness of two different machine learning algorithms based on student’s paired t-test, the unpaired t-test, and McNemar’s test.

**Project Description:** Implementing a complete pipeline of data mining for a dataset.

**Environment:** Python, WEKA, scikit-learn

**Machine Learning Aug 2015 – Dec 2015**

* Understood theoretical background of different ML algorithms.
* Implemented ML algorithms from scratch in MATLAB and Python.
* Improved the performance by using optimization methods and running time of the algorithms.

**Project description:**

* Implemented different machine learning algorithms like SVM, Naive Bayes, decision tree, and neural networks, etc. in MATLAB, Python, and C++.

**Environment**: MATLAB, Python

**Natural Language Processing Jan 2016 – May 2016**

* Gathered data from the internet and developed a code to categorize them.
* Built models for each category.

**Project description:**

* Implemented different techniques for text analyzing.
* Built language model for politician and non-politician speeches.
* Developed two different models based on two different text datasets in word level. Then, I classified new speeches and prepare the report.
* Measured the performance and preparing a final report based on results.

**Environment**: Python

**EDUCATION**

* **Ph.D. of Computer Science,** University of North Texas **Apr 2015 – March 2020**

Ph.D. dissertation focuses on Machine Learning & Data Privacy

* **Research Visitor,** University of Rutgers (DIMACS) **Summer 2016**
* **Master of Mathematics,** Institute for Advanced Studies in Basic Sciences **June 2008**
* **Bachelor of Applied Mathematics,** Amir-Kabir University of Technology **July 2005**

**TECHNICAL SKILLS**

* **Operating Systems:** UNIX, LINUX, and Windows 10/8/7/XP.
* **Databases:** MS SQL Server 2008, and MySQL.
* **Database Base Languages:** SQL.
* **Programing Language:** Python, C, and C++.
* **Machine Learning Packages:** Keras, TensorFlow, PyTorch, scikit-learn, WEKA, Matlab, SAS, SASPY.
* **Distributed Computing:** Hadoop, Spark.

**HONORS AND AWARDS**

* Student Travel Grant: CCS 2015, ACSAC 2016, CCS 2017, ACSAC 2018.
* Best paper nominated, best poster award, CODASPY, 2019

**PUBLICATIONS**

* J. H., Jafarian, and H. Takabi, and H. Touati, and **E. Hesamifard**, and M. Mohamed. Towards a General Framework for Optimal Role Mining: A Constraint Satisfaction Approach. SACMAT ’15, 2015.
* **E. Hesamifard**, and H. Takabi, and M. Ghasemi. CryptoDL: Towards Deep Learning over Encrypted Data. ACSAC 2016. December 2016, Los Angeles, California, USA, 2016.
* H. Takabi, and **E. Hesamifard**, and M. Ghasemi. Privacy-Preserving Multi-Party Machine Learning with Homomorphic Encryption. PMPML16, NIPS 2016 Workshop, Barcelona, December 9.
* **E. Hesamifard**, and H. Takabi, and M. Ghasemi. Privacy-preserving Machine Learning in Cloud. CCSW17, 9th ACM Cloud Computing Security Workshop. Dallas, Texas, USA.
* **E. Hesamifard**, and H. Takabi, and M. Ghasemi. Privacy-preserving Machine Learning as a Service. PoPETs, 2018, issue 3.
* **E. Hesamifard**, and H. Takabi, and M. Ghasemi. Deep Neural Networks, Classification over Encrypted Data. CODASPY, 2019. **Nominated for the best paper award; Received best poster award.**