

Aaron Reich

782 Peachtree St NE, Atlanta, GA 30308 | agr505@gmail.com | areich8@gatech.edu | 863-450-7132

Table of Contents

Education, Publications, Experience	1
Natural Language Processing Machine Learning Engineer	1
Experience (Continued).....	2,3
Natural Language Processing Machine Learning Engineer (Continued)	2
Software Engineer.....	2
Research Assistant	2
Skills	3
Graduate School Projects, Homeworks, and Coursework	3,4
Work Related Projects.....	4,5,6
Undergraduate School Projects	6,7

Education

MASTER OF SCIENCE IN COMPUTER SCIENCE, SPECIALIZATION IN MACHINE LEARNING (IN PROGRESS)| AUGUST 2019 TO PRESENT | GEORGIA INSTITUTE OF TECHNOLOGY (ON CAMPUS - ATLANTA, GA) | OVERALL GPA: 4.0

BACHELOR OF SCIENCE IN COMPUTER SCIENCE | AUGUST 2017 | FLORIDA ATLANTIC UNIVERSITY

- Overall GPA: 3.5, Dean's List, and Graduated Cum Laude
- Co-Founder and President of Upsilon Pi Epsilon International Honor Society for the Computing and Information Disciplines Chapter at FAU, Awarded the Faculty Award for Outstanding Leadership

Publications

- Yang, K., Nam, K.W., Qutbuddin, A. et al. Size constrained k simple polygons. *Geoinformatica* (2020).
<https://link.springer.com/article/10.1007%2Fs10707-020-00416-9>
- Reich, Aaron, Roxana Ohriniuc, and KwangSoo Yang: Size Constrained K Simple Polygons. In Proceedings of 26th ACM SIGSPATIAL Intl. Conference on Advances in Geographic Information Systems, (2018)
<https://dl.acm.org/doi/10.1145/3274895.3274962>
- Ohriniuc, Roxana, Aaron Reich, and KwangSoo Yang: Coverage Constrained Spatial CO-clustering. In Proceedings of 26th ACM SIGSPATIAL Intl. Conference on Advances in Geographic Information Systems, (2018)
<https://dl.acm.org/doi/10.1145/3274895.3274960>

Experience

NATURAL LANGUAGE PROCESSING MACHINE LEARNING ENGINEER | PIONETECHS, CONTRACTOR FOR CERNER| JULY 2020 – PRESENT

- Responsible for the investigation of a Text Classification Machine Learning approach for the evaluation of outcomes resulting from applicant credentials validation automation processes. Identified limitations of the current Expert system and the possible improvement in accuracy and generalization ability a Text Classification Machine Learning approach

provides.

- Responsible for the planning of and conducting of an experiment to compare the performance of the current expert system and NLP ML models. Designing of the experimental small dataset with limited data to be an accurate sample of the distribution of outcome text from portals that credentials validation orders are placed for.
- Built an inline Javascript and CSS parser to clean out inline CSS and Javascript that was in the text and removing symbols from the text preparing clean text for the ML model. This had to be created due the DOM and script tags no longer being in place in the text making the use of available tools like htmlagilitypack or beautifulsoup not able to be used. Responsible for designing ways to reduce text to only relevant text for the text classification problem since extracting of the web page text often extracts irrelevant information.

SOFTWARE ENGINEER | PIONETECHS, CONTRACTOR FOR CERNER| MAY 2017 – PRESENT

- While working for Pionetechs, I work as contractor for Cerner developing and supporting Cerner's HealtheHistory. It retrieves patient records in near real-time from accessible patient portals. This electronic delivery enables quicker decision making for law firms and life insurance providers such as New York Life. I develop Visual Basic .NET classes to direct and validate the automation processes and develop associated IMacros scripts. This supports the record retrieval functionality of HealtheHistory. I work with Cerner Operations to ensure orders are completed utilizing HealtheHistory's automation processes.
- I am responsible for maximizing automation process coverage of healthcare provider portals that are receiving applicant credentials validation and applicant record retrieval orders from clients. Development time for different automation processes are taken into consideration when developing strategies for future work efforts. I am also responsible for increasing the accuracy of the outcomes resulting from applicant credential validation automation processes by the developing the expert system to most accurately return outcomes based on hardcoded rules for what web page text contains.
- In order to decrease the amount of development of automation classes involved with the supporting of the automation processes of HealtheHistory, I am responsible for the development of projects with the objective of increasing code reuse where applicable and abstracting the automation processes in a way to allow for non-developers to be able to configure applicable automation processes for patient portals from an internal website. These projects involve extensions to the internal website, automation code library, and automation development application.
- Specific projects include an automation design tool extension to the internal website to be used for dynamically generating automation processes involved with patient record retrieval from specific types of patient portals. I developed the database, backend, and frontend designs for it and implemented them. I designed and implemented the extension to the automation class library consisting of an automation class that can be dynamically configured from the configurations entered on the extension to the internal website. I designed and implemented the extension to the automation development application so that way dynamic automation configurations could be tested before being configured on the internal website.

RESEARCH ASSISTANT | FOR DR. KWANGSOO YANG | FEBRUARY 2018 – APRIL 2020

- Involved in the narrowing down of the research topic from surveying current work. Involved in the formulation of the Nearest Neighbor Triangulation and Merging algorithm published in "Size Constrained k Simple Polygons" which was in the conference proceedings of ACM SIGSPATIAL in 2018. Involved in the testing of the algorithm for all theoretical cases of polygonal shapes. Involved its Java implementation. Involved in the formulation of the Bipartite Space Shrinking algorithm published in "Coverage Constrained Spatial CO-clustering" which was in the conference proceedings of ACM SIGSPATIAL in 2018.
- Involved in the reformulation of the Nearest Neighbor Triangulation and Merging algorithm with a novel approach outperforming it called the "Nearest Neighbor Point and Merging" algorithm. This work was published in the Geoinformatica Journal.

INFORMATION TECHNOLOGY INTERN | POLEN CAPITAL MANAGEMENT | MAY 2016 – JULY 2016

- Developed applications as well as database design for the building of a data warehouse to enable automation in the reporting of business metrics to executive management. Also created custom SharePoint Add-in for the viewing of the Distribution team's completed tasks.

WEB CONSULTANT | D-TECH BUSINESS SOLUTIONS – A XEROX AUTHORIZED AGENCY | FEBRUARY 2016-JUNE 2016

- In charge of the managing of the company's website, increasing functionality, adding content, and magnifying SEO for it

WEB DEVELOPMENT INTERN | BRICKS & MORTAR CREATIVE WEB AGENCY| MAY 2015- AUGUST 2015

- Worked on JS frameworks including Node.JS, Express.JS, WordPress, and utilized MongoDB

Skills

- Fluent in: Python, C, C++, C#, Java, Visual Basic .NET
- Proficient with: Numpy, Scikit-learn, Pytorch, Pandas, JavaScript, MATLAB, Microsoft SQL Server, MySQL, MongoDB, PostgreSQL, LaTeX
- Experience with the following: Jupyter Notebook, Visual Studio, Eclipse, Netbeans, Google Colab, Google Cloud, Github, Overleaf

Graduate School Projects, Homeworks, and Coursework

COURSE PROJECT FOR WEB SEARCH AND TEXT MINING: INSINCERE AND TOXIC QUESTION DETECTION

- Conducted a Natural Language Processing Machine Learning experiment for the prediction of the intention of questions asked on Quora as being either sincere or insincere. Performed a literature survey. Performed dataset preparation and the calculation of raw data statistics. Performed data analysis including of the top 20 uni-grams, bi-grams, and tri-grams, and the Ch-Squared statistic of the tokens in the corpus.
- Trained two models to serve as baselines. The first of which was an Attention-based Bi-LSTM Recurrent Neural Network. The other baseline model was a Support Vector Machine utilizing a Linear Kernel. Both models used a pre-trained GloVe embeddings. SVM used a bag-of-centroids representation for each Quora question. Bi-LSTM outperformed the SVM model in terms of F1-scoring and precision whereas the SVM model performed better in terms of recall. With the completion of these two baseline models, we proceeded to evaluate our proposed models which utilized BERT and DistilBert, and they featured improved performance in the areas of precision and recall. Also experimented with cased text and uncased text's affect on the models performances.
- Tools Used: Jupyter Notebook, Python, PyTorch, Numpy, Scikit-Learn, Colab
<https://github.com/agr505/Insincere-and-Toxic-Question-Detection/>

WEB SEARCH AND TEXT MINING (CSE 6240) HOMEWORKS AND COURSEWORK

- Homework 1: Cleaned a text dataset, created design matrices (word count matrix, binary word matrix, class imbalanced word count matrix, and tfidf matrix) and performed a classification experiment with linear SVM and logistic regression and interpreted the results.
- Homework 2: Trained Word2Vec word embeddings and used pre-trained Word2Vec word embeddings to perform a classification experiment with Random Forest. KMeans clustering was used to create a bag-of-centroids representation for each document in the corpus
- Homework 3: Used Snap.py for working with graphs/networks to experiment with Information-based Cascading for a simulated election. Implemented the effects of voters on the social graph during a decision period. Implemented the effects of advertising and campaign events on the results of the election. Interpreted the degree distributions of the two graphs used and their effects on the election results.
- Coursework: Text Mining, NLP, Network Science, Graph Machine Learning, Knowledge Graphs, Recommender Systems

MACHINE LEARNING (CS7641) HOMEWORKS AND COURSEWORK

- Implemented K-Means, Gaussian Mixture Model, Linear Regression, Ridge Regression, PCA, Decision Tree, and Random Forest. Covered hierarchical and density-based clustering algorithms as well as unsupervised evaluation metrics such as the Jaccard coefficient, Rand Statistic, Fowlkes-Mallows Measure, Beta-CV measure, Normalized Cut, Silhouette Coefficient, Davies-Bouldin Index.

- Went over the Karush–Kuhn–Tucker conditions in ML algorithms as well as the geometric interpretation of the optimization of Ridge and Lasso Regression. Worked on kernel maps for SVM. Covered the Expectation-Maximization algorithm as well during study of K-Means and Gaussian Mixture Model. Covered Naive Bayes and Logistic Regression.

COURSE PROJECT FOR MACHINE LEARNING: MOVIE REVENUE PREDICTION

- Performed dataset analysis, pre-processing, and feature reduction (PCA and XGBRegressor) on a (TMDB) 5000 Movie Dataset. Performed Linear Ridge Regression to predict the revenue of a movie. Binned the revenue amounts into bins in order to perform classification of revenue intervals by classifiers (SVM and Random Forest). Analyzed results with scatter plots and confusion matrices.
- Tools Used: Python, Numpy, Scikit-Learn

<https://github.com/agr505/MovieRevenuePredictionMachineLearningProject>

Work-Related Projects for Pionetechs and Cerner

TEXT CLASSIFICATION MACHINE LEARNING APPROACH FOR THE EVALUATION OF OUTCOMES

- Responsible for the investigation of a Text Classification Machine Learning approach for the evaluation of outcomes resulting from applicant credentials validation automation processes. Identified limitations of the current Expert system and the possible improvement in accuracy and generalization ability a Text Classification Machine Learning approach provides.
- Responsible for the planning of and conducting of an experiment to compare the performance of the current expert system and NLP ML models. Designing of the experimental small dataset with limited data to be an accurate sample of the distribution of outcome text from portals that credentials validation orders are placed for.

INLINE JAVASCRIPT AND CSS PARSER AND LOG DATA TRANSFORMER

- Built a log data transformer to cycle through log files parsing each one to retrieve the extracted webpage text and associated outcome returned by the expert system. The data is then stored in an Excel file.
- Built an inline Javascript and CSS parser to clean out inline CSS and Javascript that was in the text and removing symbols from the text preparing clean text for the ML model. This had to be created due the DOM and script tags no longer being in place in the text making the use of available tools like htmlagilitypack or beautifulsoup not able to be used.
Responsible for designing ways to reduce text to only relevant text for the text classification problem since extracting of the web page text often extracts irrelevant information.

EPIC PORTALS FILE ACQUISITION AUTOMATION DESIGN TOOL

- Designed and implemented database, backend, and frontend aspects of the Epic portals File Acquisition automation Design Tool, consisting of an extension to Data Source Admin Repo MVC ASP.NET site, an extension to the automation class library, and an extension to the Automation Dev Application. The extension to the Data Source Admin Repo allows for the configuring of automation processes for future Epic patient portals from Data Source Admin Repo. Implemented a backend combinatorial algorithm for finding the available ordering numbers that a user can configure new automation blocks for. The output of the algorithm is used to populate a dropdown in the design tool.
- The extension to the automation class library dynamically generates IMacros scripts and utilize appropriate automation process validation code based on the metadata and automation process design performed in the extension to the Data Source Admin Repo UI. The automation process validation code used in this extension is common functionality among all of the Epic automation classes which were designated for specific Epic patient portals. The extension to the automation class library allows for JSON data storing the common elements of IMacros scripts to be deserialized and

test data representing Epic patient portal HTML metadata from a user of Data Source Admin Repo to be used to dynamically build iMacros scripts. The order of the executed iMacros scripts is defined by the user and test data incorporated different orderings to test the dynamic execution.

- The extension to the Automation Dev Application replicates the functionality in the Data Source Admin Repo ASP.NET site in a Windows Forms Application allowing for developers to test automation designs made in the website.
- Tools Used: Visual Basic .NET, MS Visual Studio 2017, iMacros, HTML, JSON

DYNAMIC VALIDATION AUTOMATION EXTENSION

- This extension consisted of development to the Data Source Repo Admin website, Automation Development Application, and the patient credentials validation automation class responsible for generic patient credentials validation across all applicable patient portals. This allowed for the input of metadata of HTML frames in Data Source Admin Repo enhancing the capabilities of the Visual Basic .NET class involved with the dynamic generation of iMacros scripts. Developed the Automation Development Application to have this extended functionality as well. Performed code reviews and tested new code to be deployed for the Data Source Repo Admin website.
- Tools Used: JavaScript, Visual Basic .NET, Microsoft Visual Studio 2017, Microsoft SQL Server

EPIC ACQUISITION BASE ABSTRACT CLASS

- Designed and implemented Epic Acquisition Base Abstract Class in order to create reusable code to generically handle common functionality utilized in automation classes for EPIC EMR portals which greatly reduced development time for these automation classes. Tools Used: Visual Basic .NET, Microsoft Visual Studio 2017, XPath, HTMLAgilityPack

DEVELOPMENT OF AUTOMATION CLASS CODE LIBRARY

- Developed Visual Basic .NET classes to direct and validate the automation processes and developed associated iMacros scripts. This supports the patient credentials validation and the patient data retrieval functionality of HealtheHistory. The patient data retrieval functionality incorporated in the automation processes utilizes algorithms I implemented to allow for the portals' different HTML tables to be traversed for the acquisition of the needed documents. HTMLAgilityPack and XPath are utilized to count number of rows in a table. Downloaded files are encrypted and placed in directories with specific naming conventions. Every phase of the algorithms is validated to ensure correct behavior. Visual Basic .NET code was developed for the directing of the algorithms' logic and validation of behavior.
- Developed an algorithm for a particular portal involving the extraction of each page of the HTML table, parsing of the HTML rows, cleaning of the data, and then populating a CSV file with the data. iMacros scripts were developed for interaction with websites. Wrote SQL queries in order to perform various operations on iMacros scripts and information associated with iMacros scripts located in the database.
- Developed the SQL install scripts for the installation of the iMacros scripts on the previous system. Currently work with provider configuration JSON files holding the provider data and iMacros scripts which the new system uses for a more streamlined workflow of installing iMacros scripts in the database
- Developed automation classes for patient portals in which two steps are needed to validate patient credentials. HTMLAgilityPack and XPath are utilized for the parsing of HTML metadata to then be presented to the user in the HealtheHistory UI.
- Developed the Visual Basic .NET class involved with the dynamic generation of iMacros scripts in order to enable patient credentials validation with more healthcare portals.
- Tools Used: Visual Basic .NET, SQL, Microsoft SQL Server, Microsoft Visual Studio 2015, iMacros, XPath, HTMLAgilityPack, HTML, iMacros, JSON

Work-Related Project for Collateral Opportunities

VERSABADGE REGISTRATION ANDROID APP

- Developed Android App allowing users to scan QR codes located on medical supplies in hospitals and doctors' badges and parses QR code data to confirm the data is in the correct format. It uses a SOAP API to see if the asset is already in the database and ensures that the asset is of an existing asset type. If it is it allows its information to be updated and if it's not it will allow for the adding of a new asset. Utilizes REST API as well for acquiring associated organizations. All network operations are executed using AsyncTasks in order to not have them executed on the UI thread. Built a password protected Settings page for the App that has persistence for the default settings. Integrated the QR scanner library for use within project. The app greatly decreased the time involved in the registration of hospital supplies and doctors to be then tracked by VersaBadge.
- Tools Used: Java, SOAP API, REST API, Android Studio, Gradle, Github

Work-Related Projects for Polen Capital Management

POLEN CAPITAL MANAGEMENT EXCEL TO SQL DATABASE APPLICATION

- Developed a Windows Forms Application to transfer data from various Excel files of different formats holding company data to a SQL Server database. Created tables in SQL Server Management Studio 2008 and mapped them in the application using LINQ for native data querying. Will send message if authentication for using the database has failed.
- Tools Used: C#, Transact-SQL, .NET (including LINQ and MS Office Interop Frameworks), MS Visual Studio 2015, SQL Server Management Studio 2008

<https://github.com/agr505/PCMExceltoSQLDatabase>

POLEN CAPITAL MANAGEMENT DYNAMIC EXCEL TO SQL DATABASE APPLICATION

- Developed a Windows Forms Application to transfer data from any Excel file to a SQL database. Utilizing SQL Command objects, any Excel file's data can be transferred once the various data types have been specified. Will send message for user to complete the application's first tab if database has not been specified or does not exist.
- Tools Used: C#, Transact-SQL, .NET (including SQL Client and MS Office Interop Frameworks), MS Visual Studio 2015, SQL Server Management Studio 2008

<https://github.com/agr505/DynamicExceltoSQLDatabase>

MONTHLY REPORTS SHAREPOINT ADD-IN

- Created custom SharePoint Add-in for the viewing of the Distribution team's completed tasks. Custom Column types for the months the task was completed in was added to the SharePoint List. Wrote JavaScript for the custom rendering of the completed task columns dependent on the current date.
- Tools Used: JavaScript, XML, SharePoint Office 365 Online, MS Visual Studio

<https://github.com/agr505/MonthlyReportsSharePointAdd-in>

Undergraduate School Projects

FLIGHT ASSISTANT DATABASE APPLICATION

- A Windows Form Application that utilizes a MySQL database for obtaining and storing flight information. There are five database tables: Flights, Airline, Arrive, Departs, and Airport. They have important relationships to each other and embody normalization concepts. The application itself creates queries based on the specific criteria the user enters. This project was created by a partner and myself for my Applied Database Systems (COP4703) elective course.
- Tools Used: C#/.NET (including MySqlClient Frameworks), SQL, MySQL Workbench, Microsoft Visual Studio 2015

<https://github.com/agr505/Flight Assistant>

REAL ESTATE FINDER JAVA SWING APPLICATION

- A Java Swing Application allowing sellers to post real estate listings and customers to browse listings and add properties they are interested in to their favorites. Customers can also submit their contact information to the seller if they are interested in buying the property. This application utilizes a rich MVC architecture along with utilizing the Strategy design pattern to create the Property Description Page based on the customers actions and recreate the page in real time. This application utilizes serialization for persistence. The application also uses the Observer and Iterator design patterns. This project was created by myself and two other partners for my Object-Oriented Design and Programming elective course (COP4331). The design for the application includes Use Cases, CRC diagram, UML Class diagrams, Sequence diagrams, and State diagrams. I did most of the application's design and about 75% of the code.
- Tools Used: Java (including Iterator, LinkedList, Serializable, Swing frameworks), NetBeans

<https://github.com/agr505/RealEstateFinder>

OBJECT-ORIENTED DESIGN AND PROGRAMMING (JAVA) HOMEWORK REPOSITORY

- This is my repository for Object-Oriented Design and Programming (COP 4331) homework
- Tools Used: Java, Netbeans

<https://github.com/agr505/ObjectOrientedHomework>

MATLAB MACHINE LEARNING CAT OR DOG IMAGE CLASSIFIER PROJECT

- Developed a machine learning project for determining if an image is a picture of a dog or of a cat. 1,170 images from the Kaggle data set were used for feature selection performed by a pre-trained Convolutional Neural Network to create a Cosine K-Nearest Neighbor classifier and a Cubic Support Vector Machine classifier. 25,000 images were then used for a second feature selection to create another Cubic Support Vector Machine since it was the better performing classifier. The entire machine learning workflow was performed including partitioning the data, feature selection, classifier selection, training, cross-validation and testing. The MATLAB Classification Learner App was used to help create the classifiers. This project was performed by two partners and myself for my Introduction to Artificial Intelligence elective course (CAP4630).

<https://github.com/agr505/CatorDogMachineLearningImageClassifierProject>

CPU SCHEDULER SIMULATION

- Developed three CPU scheduling algorithms for simulating how computer operating systems schedule processes. I implemented FCFS (First Come First Serve) non-preemptive, SJF (Shortest Job First) non-preemptive, and MLFQ (Multi Level Feedback Queue) preemptive.
- After processes complete a CPU burst, they then go in I/O and once they complete I/O they are scheduled by the algorithm. Average wait time, turnaround time, CPU utilization, and response time for each algorithm are calculated.
- The MLFQ implementation embodies object-oriented principles by containing a class MLFQ_Queue which inherits from the generic Queue<T> and each MLFQ_Queue contains references to other MLFQ_Queue's relative to their respective priorities. The two higher priority queue's use Round Robin with different time quanta for their own scheduling while the lowest priority queue uses FCFS for scheduling
- Tools Used: C#/ .Net, Microsoft Visual Studio 2015

<https://github.com/agr505/CPU-Scheduler>

SORTING ALGORITHMS IMPLEMENTATION AND RUN-TIME ANALYSIS

- Implemented Insertion Sort, Quick Sort, and Heap-Sort algorithms in Java and computed and graphed the running times for each algorithm with various input sizes and compared them to the theoretical running times using Big O-notation.

<https://github.com/agr505/Sorting-Algorithms>

