Master of Science in Applied Data Science

Syracuse University

IST 782 | Applied Data Science Portfolio

Matthew L. Pergolski

Table of Contents

Introduction

In this report, I am thrilled to present my journey and achievements as a student in the Master of Science in Applied Data Science program at Syracuse University. Throughout my time in the program, I have developed a deep passion for solving complex problems through the power of data.

The rapidly evolving field of data science has transformed industries, improved decision-making processes, and created positive impact on society. It is precisely why I wanted to become associated with the field and make it a career path. Recognizing the need to equip myself with cutting-edge knowledge and skills, I entered this program to further my expertise and contribute to the broader field in industry.

The comprehensive curriculum of the Master of Science in Applied Data Science program – which covers data science topics from data collection and preprocessing to advanced machine learning and big data analytics – has provided me with a solid foundation to tackle real-world challenges. The program's emphasis on practical application and problem-solving has been particularly valuable, as it has enabled me to apply theoretical concepts to real-world scenarios.

Throughout this report, I will highlight projects, coursework, and experiences that have shaped my journey in the program. I will highlight how the program's objectives and outcomes have aligned with my own goals and how I have leveraged the knowledge and skills gained to enter the field via Lockheed Martin – a leading aerospace and defense company.

The report will begin by discussing my personal interest in the program and how it aligns with my career aspirations. I will then highlight specific courses that have been instrumental in my growth on my data science journey. These include the IST 736 “Text Mining” project, the IST 718 “Big Data Analytics” project, and the IST 691 “Deep Learning in Practice” project – each of which demonstrates my ability to apply advanced data science techniques to real-world problems.

I will also discuss the data science lifecycle and how the program has equipped me with the skills and knowledge necessary to navigate each stage of the process effectively. Additionally, I will highlight the program's objectives and outcomes and provide evidence of how I have achieved these goals through my coursework and projects.

Throughout the report, I will emphasize the importance of effective communication, both in terms of presenting technical findings to stakeholders and collaborating with cross-functional teams, something that is crucial to execute well in industry. I will also discuss

the ethical considerations that are crucial in the field of data science and how the program has addresses these.

As I reflect on my journey through the Master of Science in Applied Data Science program, I am grateful for the opportunities it has provided me to grow both personally and professionally. I am excited to leverage the skills, knowledge, and network I have gained to make an impact in the world of data science.

Personal Interest in Data Science Program

When I began my professional career in 2018, my work primarily revolved around using Microsoft Excel for various tasks. Initially, I was excited about learning the diverse range of formulas and functions that Excel offered, and I even took the initiative to learn many functions of Visual Basic for Applications (VBA) to automate specific tasks; however, as time progressed, I started to realize that a significant portion of my work felt repetitive, as I found myself performing the same tasks repeatedly in similar contexts. This realization sparked me to contemplate whether a more effective approach existed to tackle these tasks – perhaps by leveraging more advanced tools and techniques.

Determined to enhance my skills, I dedicated myself to learning as much as possible about Excel and VBA. Nevertheless, after a certain point, it became apparent that these tools had their limitations, and I aspired to expand my horizons and develop more advanced automation workflows. It was clear that I needed to expand my skill set and acquire new knowledge to achieve these goals.

The COVID-19 pandemic in 2020 presented an unexpected opportunity for me to reflect on my career aspirations and explore new avenues for growth. With more time available, I decided to seek online resources to further my understanding of automation and programming. This exploration led me to discover the Python programming language, which I had previously heard of but had never considered learning myself. As I researched more about Python, I became increasingly interested about its potential for automation with respect to my current scope of work.

Eager to grasp the fundamentals of Python, I enrolled in online courses and worked through various tutorials. Despite the initial challenges, I found the learning process to be engaging. With each new concept I understood, I could envision the possibilities of applying Python in my current career, which furthered my motivation to continue learning.

As my knowledge in this space expanded, I began to contemplate how I could transform this passion into a career change. I recognized that I wanted to progress beyond using Excel and start working with more powerful tools. It was during this time that I discovered Syracuse University's Master of Science in Applied Data Science program, which immediately captured my interest. The program's curriculum appeared to encompass all

the essential aspects of data science – from data collection and preprocessing to machine learning and data analytics, making it an ideal fit for myself. At the time – and still to this day – I explored educational paths in data science since it heavily dealt with Python and creating pipelines to automate tasks.

Program Highlights and Appeal

As I explored the possibility of joining the program, one of the most appealing aspects of the MS degree was its emphasis on practical, real-world applications of data science, machine learning, and artificial intelligence. Rather than focusing solely on theoretical concepts, the program aimed to provide hands-on experience in utilizing data to solve concrete problems. This approach aligned with my desire to gain practical skills that I could apply in my professional experiences.

Pursuing a Master's degree is a significant financial investment, which led me to explore the possibility of securing financial support from my employer. Many companies offer tuition assistance programs to encourage employees to further their education and progress their careers, and I believed that if I could demonstrate how the data science program would directly benefit my work, I may be able to obtain financial aid to offset the costs.

The idea of applying to Syracuse's program filled me with excitement – as it represented the ideal next step in my career journey. By building upon my existing skills and interests while opening up new opportunities for the future, I believed that this program would provide me with the knowledge and experience necessary to break into the field of data science. Although I recognized that the path ahead would be challenging, I felt prepared and motivated for the journey ahead.

Career Goals and Aspirations

My career goals and aspirations in the field of data science are centered around mastering both the fundamental and advanced topics of this rapidly evolving discipline. As someone who possesses a deep passion for lifelong learning, I am committed to continuously expanding my knowledge and skills in data science through various avenues. These include online learning platforms such as Coursera, LinkedIn Learning, and Pluralsight, as well as engaging in personal research projects to stay at the forefront of the latest developments in the field.

About halfway through my data science program at Syracuse University, I was fortunate enough to be offered a position as a Data Scientist Analyst, Level II, at my current company – which later evolved into a Level III position. This opportunity came as a testament to the

skills and knowledge I had acquired thus far in my academic journey, and I was thrilled to begin gaining practical industry experience while simultaneously continuing to build my knowledge through my coursework.

The combination of hands-on experience in a professional setting and the academic training provided by the Master's program has been invaluable in shaping my career trajectory. It has allowed me to apply the theoretical concepts learned in the classroom to real-world scenarios – further reinforcing my understanding of data science fundamentals.

As I progress in my career, my primary goal is to continue mastering as many data science topics as possible and to effectively apply them within industry. I am dedicated to staying updated with the latest advancements in the field, as I believe that this is essential for driving innovation and solving complex business problems. By leveraging my growing skill set and industry experience, I aim to make significant contributions to my organization and to the broader data science community.

Program Objectives and Outcomes

According to the Syracuse University Master of Science in Applied Data Science program's capstone syllabus, the key objectives and outcomes of the academic program include:

- Collect, store, and access data by identifying and leveraging applicable technologies

The program equips students with the skills necessary to identify and utilize appropriate technologies for data collection, storage, and access. This involves understanding the various data sources, such as databases, APIs, and web scraping, and cloud computing options like MinIO/S3 – and selecting the most suitable tools and techniques based on the specific requirements of the project.

- Create actionable insight across a range of contexts (e.g., societal, business, political), using data and the full data science life cycle

Throughout the program, students gain the ability to generate meaningful and actionable insights from data across various domains – including societal, business, political, and others. By applying the full data science life cycle, which encompasses data collection, preprocessing, analysis, modeling, and interpretation, students learn to extract valuable knowledge from raw data. They develop the skills to identify patterns, trends, and relationships within the data, enabling them to provide data-driven recommendations.

- Apply visualization and predictive models to help generate actionable insight

The program emphasizes the importance of data visualization and predictive modeling in generating actionable insights. Students learn to create compelling and informative visualizations that effectively communicate complex data patterns and relationships to both technical and non-technical audiences. They also acquire the skills to develop and apply predictive models – such as regression, classification, and time series analysis, to forecast future trends, identify potential risks, and optimize business strategies based on data-driven insights.

- Use programming languages such as R and Python to support the generation of actionable insight

Throughout the program, students gain insight into using programming languages like R, SQL, and Python, which are essential tools in the data science ecosystem. They ultimately learn to leverage the available libraries and frameworks like Pandas, Matplotlib, Scikit-Learn, PyTorch, and others, to perform data manipulation, statistical analysis, machine learning, and data visualization tasks.

- Communicate insights gained via visualization and analytics to a broad range of audiences (including project sponsors and technical team leads)

Effective communication is a critical skill for data scientists, as they often need to convey complex technical findings to various audiences. The program trains students to present their insights and recommendations in a clear and concise manner – tailoring communication to the specific needs of their respective audience. This includes creating visually appealing and informative presentations, writing reports, and engaging in productive discussions with project stakeholders (both technical and non-technical) to ensure that the insights derived from data are effectively translated into actionable strategies.

- Apply ethics in the development, use, and evaluation of data and predictive models (e.g., fairness, bias, transparency, privacy)

The program places a strong emphasis on the ethical considerations surrounding data science practices. Students learn to identify and mitigate potential biases in data collection and analysis, ensuring that their models and insights are fair and unbiased. They also gain an understanding of the importance of transparency in data science processes, enabling stakeholders to trust and rely on the insights discovered.

By achieving these program objectives and outcomes, graduates of the Syracuse University Master of Science in Applied Data Science program are well prepared to tackle complex data-driven challenges, generate actionable insights, and make a significant impact in their chosen fields.

Data Science Life Cycle

The data science lifecycle is a comprehensive and iterative process that includes various stages – from understanding the problem domain to delivering actionable insights. Throughout the Syracuse University Master of Science in Applied Data Science program, students gain an understanding of this lifecycle and navigate each stage effectively.

Learning the application domain is the first step in the data science lifecycle. It involves understanding the specific problem or question at hand, identifying the relevant stakeholders, and gaining domain-specific knowledge. Students in the program learn to collaborate closely with subject matter experts to grasp the nuances of the application domain, ensuring that the data science solutions developed are aligned to the goals of the organization or project requirements – which many companies communicate through a charter, project or program directive (PD).

Attention to quality is a fundamental through every stage of the data science lifecycle. Students are taught to ensure the accuracy, reliability, and integrity of data throughout the process.

Effective communication with data users is essential for the success of any data science project. Engaging with stakeholders, understand requirements, and translating complex technical concepts into simple language is a skill required to be effective in industry. This involves active listening, asking clarifying questions, and providing regular updates on the progress and findings of the project.

Visualization and presentation are critical components of the data science lifecycle, as they enable the effective communication of insights to both technical and non-technical audiences. Students learn to create visually compelling and informative representations of data, using tools and techniques such as data visualization libraries, dashboards, and interactive user interface (UI) components.

Seeing the "big picture" is an essential skill for data scientists, as it allows them to understand how their work fits into the broader context of the organization or project. Students learn to consider the strategic goals, business objectives, and potential impact of their data science solutions. By maintaining a holistic perspective, they can identify opportunities for collaboration, anticipate potential challenges, and ensure that their work delivers value and aligns with the overall direction of the organization.

Data transformation and analysis are at the core of the data science lifecycle. Students gain hands-on experience in cleaning, preprocessing, and transforming raw data into a format suitable for analysis. They learn to apply a wide range of statistical and machine learning techniques to extract meaningful patterns from the data. This involves leveraging

programming languages like R and Python, as well as utilizing various libraries and frameworks specifically designed for data manipulation and analysis.

Knowing how data can be represented is another fundamental skill for data scientists. Students learn about different data structures – such as tables, data frames, and time series – and understand how to choose the most appropriate representation based on the nature of the data and the analysis requirements. They also gain knowledge of various data storage technologies, including relational databases, NoSQL databases, and big data platforms.

An overview of this lifecycle can be viewed below:

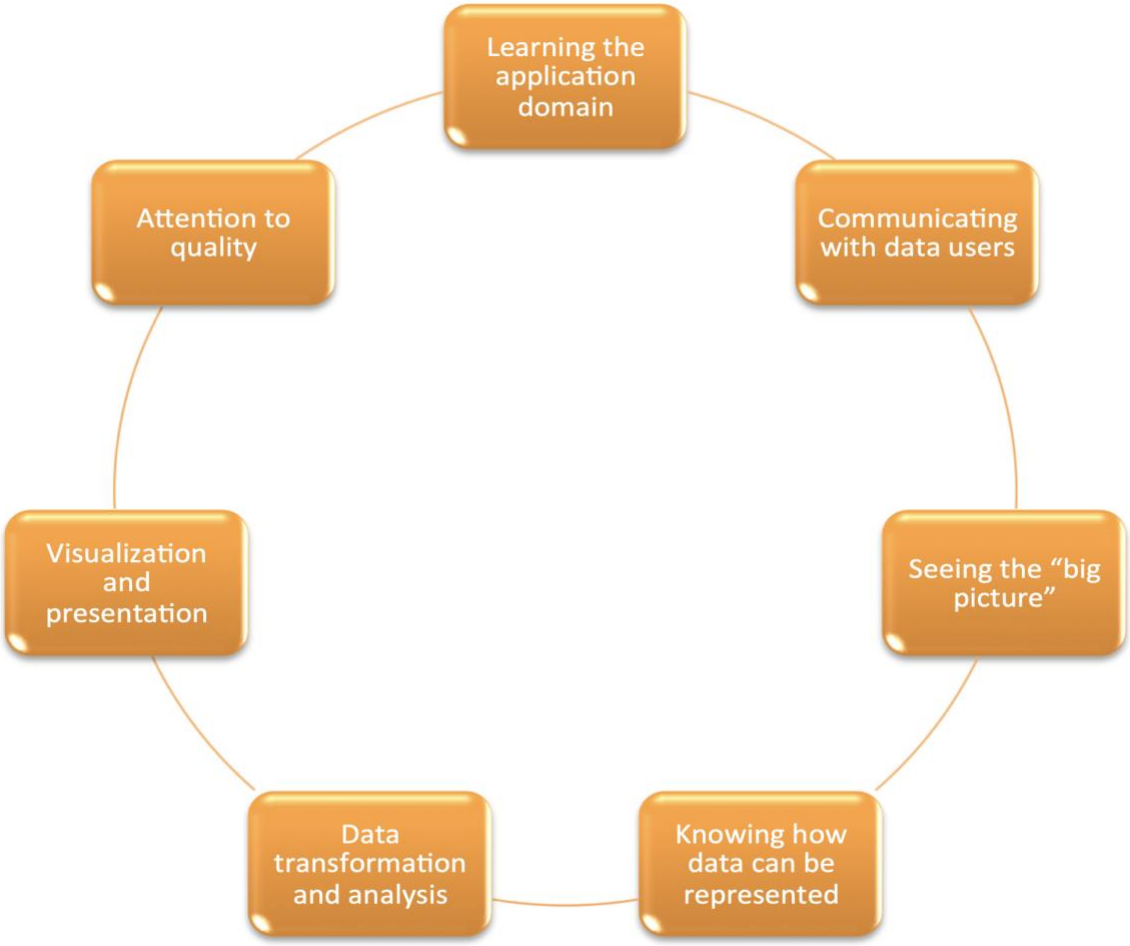

Figure 1.

Coursework and Curriculum

Throughout my journey in the Syracuse University Master of Science in Applied Data Science program, I have completed a diverse range of courses that have provided me with a comprehensive understanding of the field. The coursework has covered essential topics in data science, including statistical analysis, machine learning, big data management, and more. Below is an overview of the classes I have taken:

- IST 687 | Introduction to Data Science:

This course served as a foundation for the program, introducing the fundamental concepts and techniques in data science. It covered topics such as data collection, preprocessing, exploratory data analysis, and basic machine learning algorithms. The course provided hands-on experience with the R programming language – enabling students to gain practical skills in data manipulation and analysis.

- IST 772 | Quantitative Reasoning in Data Science:

In this course, I ventured into the mathematical and statistical foundations of data science. It covered topics such as probability theory, hypothesis testing, regression analysis, and experimental design. The course emphasized the importance of quantitative reasoning skills in data-driven decision-making and provided a solid foundation for more advanced topics in the program. The course was based on R programming.

- IST 707 | Applied Machine Learning:

This course focused on the practical application of machine learning techniques to real-world problems. It covered a wide range of supervised and unsupervised learning algorithms – including decision trees, random forests, support vector machines, and clustering methods. The course also introduced concepts like feature selection, model evaluation, and hyperparameter tuning, enabling students to build and optimize machine learning models effectively. The course was primarily R programming based.

- IST 659 | Data Administration Concepts and Database Management:

In this course, I learned about the principles and practices of data administration and database management. It covered topics such as data modeling, relational database design, SQL querying, and the basics of data warehousing. The course provided hands-on experience with database management systems like SQL server, equipping students with the skills necessary to design, implement, and maintain efficient and scalable data storage solutions.

- IST 652 | Scripting for Data Analysis:

This course focused on the use of scripting languages, particularly Python, for data analysis tasks. It covered topics such as data wrangling, data visualization, and web scraping. The course provided hands-on experience with popular Python libraries like pandas, NumPy, and Matplotlib, enabling students to efficiently analyze and visualize data.

- IST 769 | Advanced Big Data Management:

In this course, I explored the challenges and solutions associated with managing and processing large-scale datasets. It covered topics such as distributed computing, Hadoop ecosystem, Spark, and NoSQL databases. The course provided hands-on experience with big data technologies, equipping students with the skills necessary to design and implement scalable data processing pipelines.

- IST 664 | Natural Language Processing:

This course introduced the concepts and techniques of natural language processing (NLP). It covered topics such as text preprocessing, part of speech tagging, sentiment analysis, topic modeling, and other techniques. The course provided hands-on experience with NLP libraries like NLTK, enabling students to build solutions that can understand and generate insights via human language.

- SCM 651 | Business Analytics:

In this course, I learned how to apply data analytics techniques to solve business problems. It covered topics such as data-driven decision-making, predictive modeling, and optimization. The course provided hands-on experience with tools like Excel, Tableau, and R, enabling students to analyze and visualize business data effectively.

- IST 718 | Big Data Analytics:

This course focused on the application of advanced analytics techniques to big data. It covered topics such as machine learning at scale, graph analytics, and an introduction to deep learning as well as language processing. The course provided hands-on experience with data analytics platforms like Google Colab, equipping students with the skills necessary to extract insights from large and complex datasets with cloud computing hardware.

- IST 736 | Text Mining:

In this course, I learned about the techniques and applications of text mining. It covered topics such as text preprocessing, feature extraction, text classification, and topic modeling. The course provided hands-on experience with text mining libraries like gensim

and scikit-learn, enabling students to analyze and extract insights from unstructured textual data.

- IST 691 | Deep Learning In Practice:

Lastly, this course focused on the practical application of deep learning techniques to real-world problems. It covered topics such as neural networks, convolutional neural networks, recurrent neural networks, and deep reinforcement learning. The course provided hands-on experience with deep learning frameworks like TensorFlow and PyTorch, equipping students with the skills necessary to build and deploy deep learning models for various applications, such as image classification, natural language processing, and recommendation systems.

Evidence of Achieving Program Goals

Throughout my coursework in the Syracuse University Master of Science in Applied Data Science program, I have had the opportunity to apply the knowledge and skills acquired to real-world projects and assignments. In the following subsections, I will highlight three course projects in particular that I believe demonstrated my ability to achieve program goals as well as address aspects within the data science life cycle:

IST 736 | Text Mining

One notable example is the IST 736 Text Mining course, where I completed a comprehensive project that demonstrates my ability to achieve several key program goals.

The IST 736 project involved developing a text mining application that assists prospective students in exploring and gaining insights from Syracuse University's Master of Science in Applied Data Science program's course syllabi. By leveraging techniques such as term frequency analysis, document similarity search, and topic modeling, the application enables users to navigate through the vast amount of information contained within the syllabi and make informed decisions about their academic path.

This project showcases my proficiency in collecting, storing, and accessing data by identifying and leveraging applicable technologies. I successfully extracted textual data from PDF syllabi documents and preprocessed the data using techniques like tokenization, lemmatization, and stop word removal. The processed data was then stored in a structured format, allowing for efficient analysis and retrieval.

Furthermore, the project demonstrates my ability to create actionable insights using data and the full data science life cycle. By applying techniques like document similarity search and topic modeling, I generated meaningful insights that assist prospective students in

understanding the key topics and themes covered in the program's curriculum. The application's search functionality allows users to query the syllabi and discover courses most relevant to their interests, while the topic modeling component uncovers the overarching themes present within the corpus.

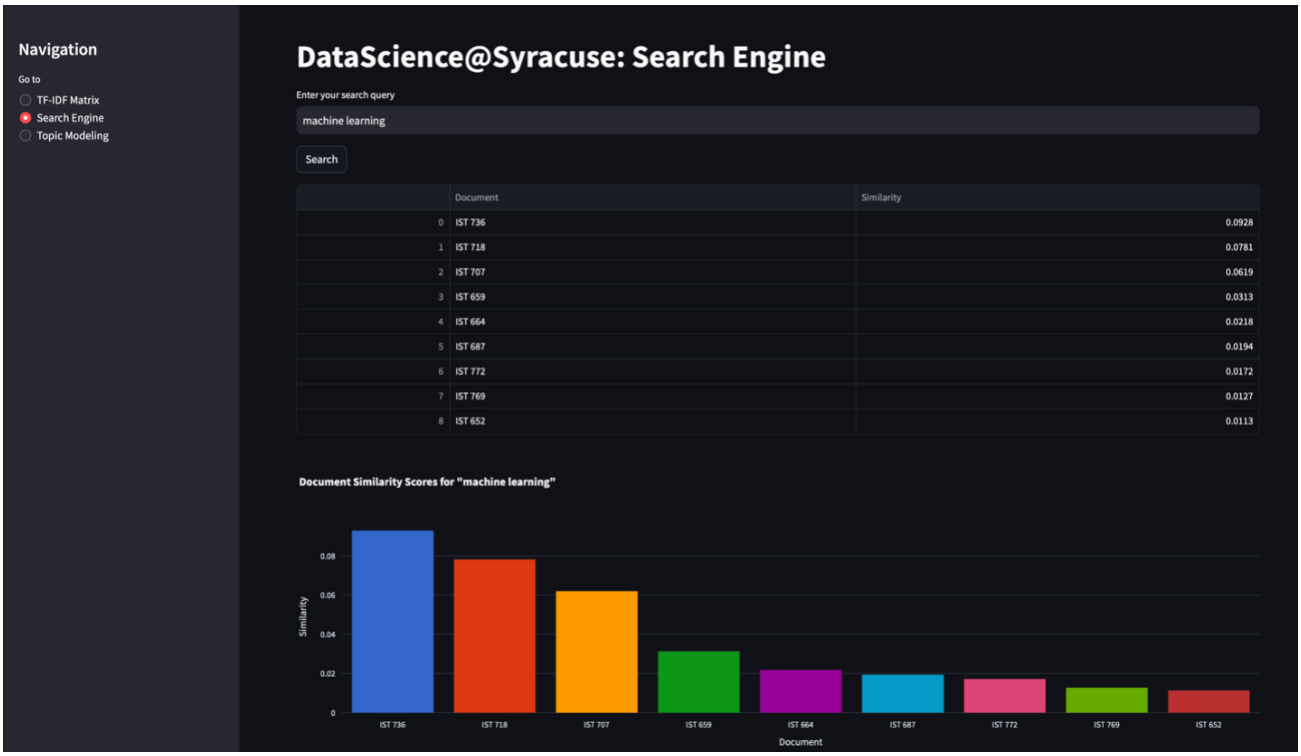

Figure 2.

The IST 736 project also highlights my proficiency in applying visualization and predictive models to generate actionable insights. The application includes interactive visualizations, such as a bar chart displaying the similarity scores of courses based on user queries and an interactive topic modeling visualization using pyLDAvis. These visualizations enable users to explore and interpret the results effectively, facilitating data-driven decision-making.

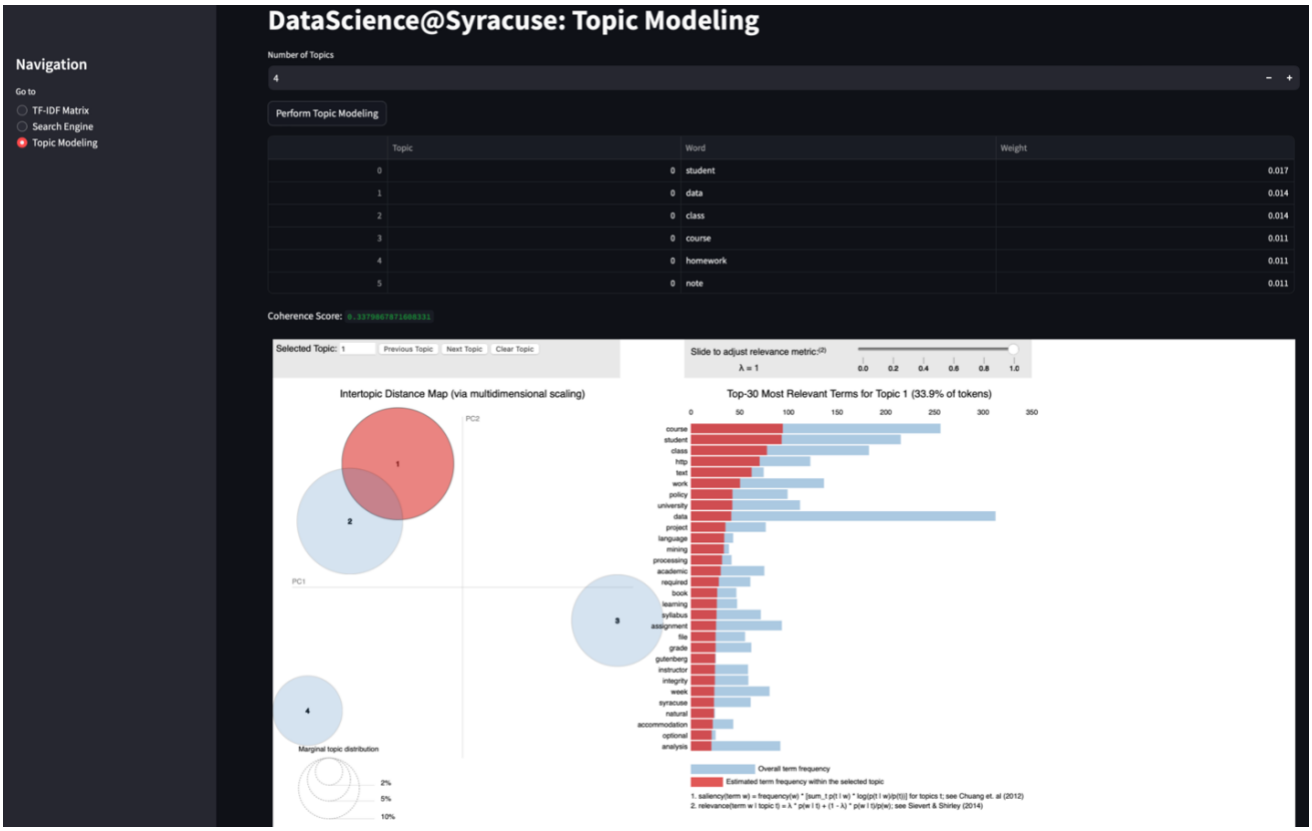

Figure 3.

Moreover, this project conveys my ability to use programming languages like Python to support the generation of actionable insights. I leveraged various Python libraries, such as pdfminer for PDF extraction, scikit-learn for document similarity calculation, and gensim for topic modeling. The application's backend was built using Python, demonstrating my proficiency in utilizing the programming language for data science tasks.

The IST 736 project also showcases my ability to communicate insights gained via visualization and analytics to a broad range of audiences. The application's user interface, developed using the Streamlit library, presents the results in a clear and intuitive manner, making it accessible to both technical and non-technical users. The visualizations and interactive components facilitate the communication of complex text mining concepts and insights to prospective students and other stakeholders.

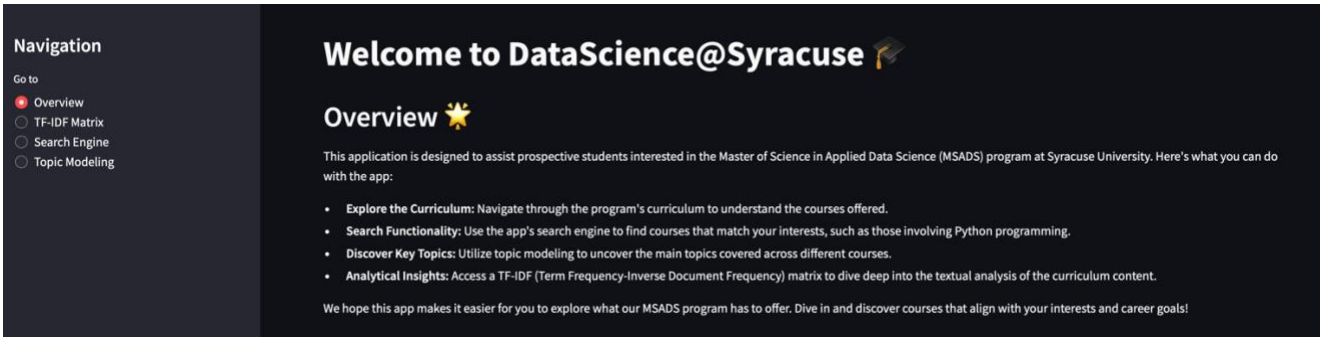

Figure 4.

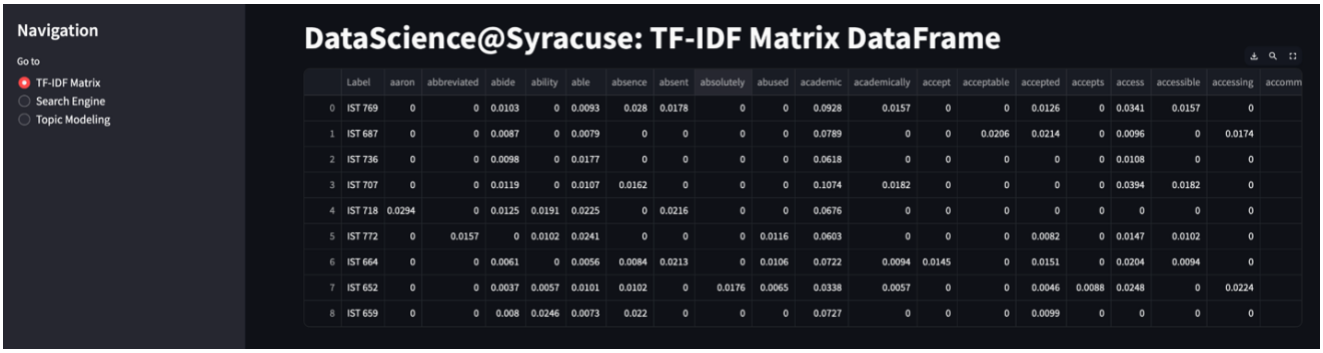

Figure 5.

Lastly, this project demonstrates my commitment to applying ethics in the development, use, and evaluation of data and predictive models. Throughout the project, I carefully considered the potential biases and limitations of the data and algorithms used. I ensured that the application provides objective information and insights without discriminating against any particular group or individual.

In conclusion, the IST 736 Text Mining course project serves as strong evidence of my ability to achieve the key objectives and outcomes outlined in the Syracuse University Master of Science in Applied Data Science program. It showcases my proficiency in data collection, storage, and access; creating actionable insights; applying visualization and predictive models; utilizing programming languages; communicating insights effectively; and adhering to ethical principles in data science practices.

IST 718 | Big Data Analytics

Another significant project that demonstrates my ability to achieve the program goals is the IST 718 Big Data Analytics course project. In this project, our group focused on developing an image classification model to distinguish between normal and pneumonia-affected chest X-rays. This project conveys my proficiency in applying advanced analytics techniques to large datasets and generating actionable insights.

The project involved working with a dataset of 5,863 chest X-rays from children aged one to five years old, with each image categorized as either normal or pneumonia. We preprocessed the images to ensure consistent size and normalization, enabling efficient analysis and modeling.

An example of images from the dataset as well as their respective classification labels can be previewed below:

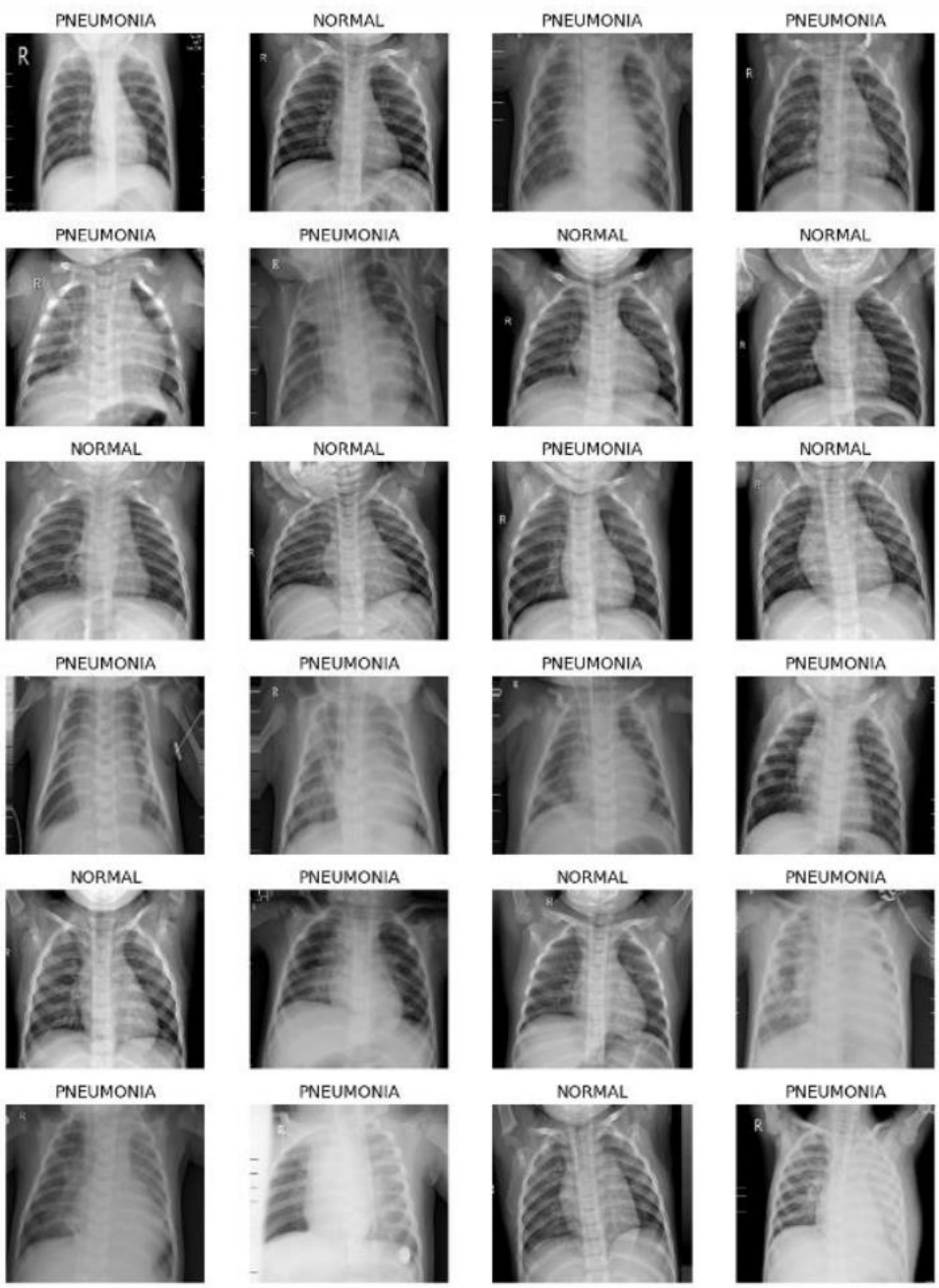

Figure 6.

We applied two deep learning models, ResNet and VGG-16, to classify the chest X-rays accurately. By leveraging the power of convolutional neural networks (CNNs), we were able to extract meaningful features from the images and train the models to distinguish between normal and pneumonia cases. This demonstrates my ability to apply advanced machine learning techniques to solve real-world problems.

An overview of the backpropagation process of the innerworkings of the models can be seen below:

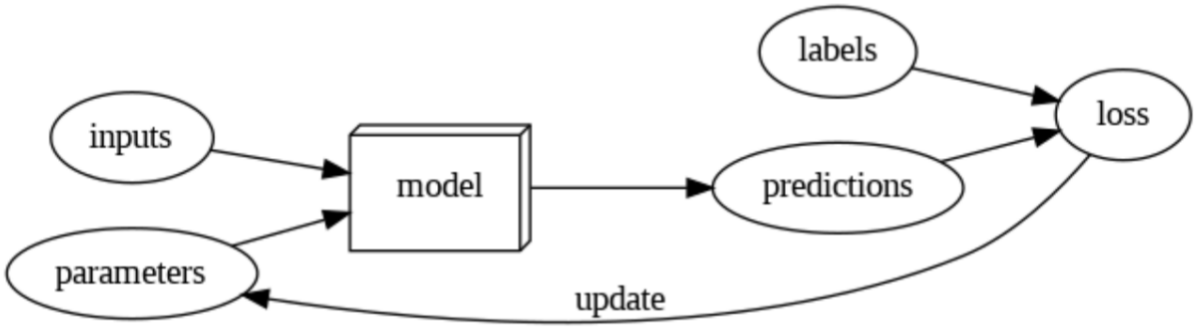

Figure 7.

The ResNet models, with varying numbers of layers (18, 50, and 152), achieved accuracies ranging from 81.89% to 85.58%. We also employed techniques like learning rate optimization to further improve the model's performance. The VGG-16 model, after fine-tuning and increasing the number of epochs, achieved an impressive training accuracy of 97.9%, but similar validation accuracies to ResNet.

To gain deeper insights into the model's decision-making process, we utilized Grad-CAM (Gradient-weighted Class Activation Mapping) to visualize the regions of the chest X-rays that the model focused on for classification. This allowed the group to interpret the model's predictions and ensure it was identifying relevant features.

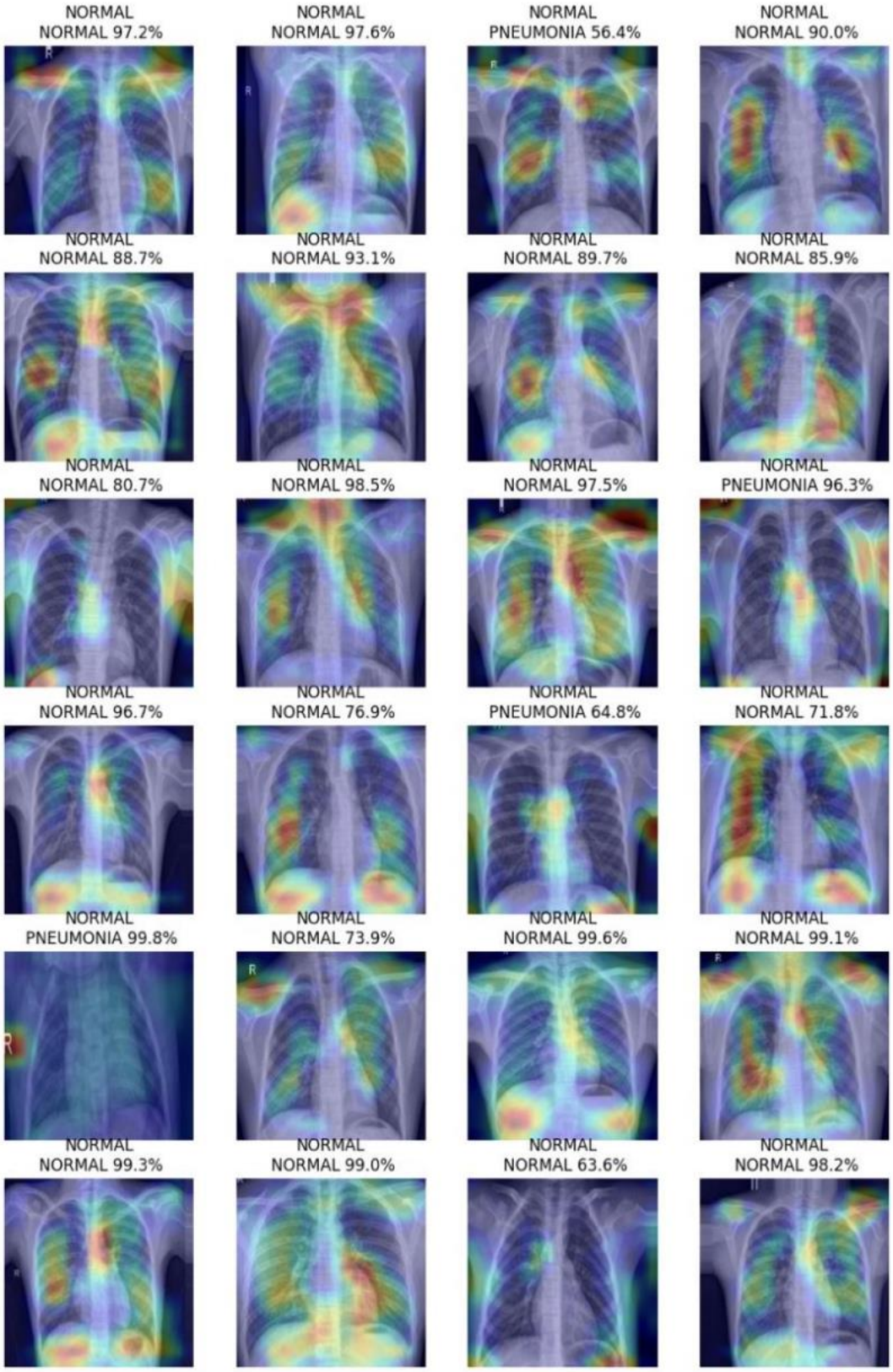
NORMAL
NORMAL 97.2%
NORMAL
NORMAL 97.6%
NORMAL
PNEUMONIA 56.4%
NORMAL
NORMAL 90.0%
NORMAL
NORMAL 88.7%
NORMAL
NORMAL 93.1%
NORMAL
NORMAL 89.7%
NORMAL
NORMAL 85.9%
NORMAL
NORMAL 80.7%
NORMAL
NORMAL 98.5%
NORMAL
NORMAL 97.5%
NORMAL
PNEUMONIA 96.3%
NORMAL
NORMAL 96.7%
NORMAL
NORMAL 76.9%
NORMAL
PNEUMONIA 64.8%
NORMAL
NORMAL 71.8%
NORMAL
PNEUMONIA 99.8%
NORMAL
NORMAL 73.9%
NORMAL
NORMAL 99.6%
NORMAL
NORMAL 99.1%
NORMAL
NORMAL 99.3%
NORMAL
NORMAL 99.0%
NORMAL
NORMAL 63.6%
NORMAL
NORMAL 98.2%

Figure 8.

Furthermore, we evaluated the VGG-16 model's performance using a Receiver Operating Characteristic (ROC) curve. The ROC curve demonstrated the model's high true positive rate and low false positive rate, with an area under the curve (AUC) close to 0.96, indicating high accuracy in classifying normal vs. pneumonia cases in some training examples.

This project highlights my (as well as the team's) ability to work with large datasets, apply advanced analytics techniques, and generate actionable insights. It demonstrates proficiency in using programming languages like Python and leveraging powerful libraries such as TensorFlow and Keras for deep learning tasks. The project also showcases the ability to communicate insights effectively through visualizations and interpret the model's decision-making process.

Overall, the IST 718 Big Data Analytics course project, along with the IST 736 Text Mining course project, provides strong evidence of the ability to achieve key objectives and outcomes outlined in the Syracuse University Master of Science in Applied Data Science program. These projects demonstrate skills in data preprocessing, applying advanced machine learning techniques, generating actionable insights, and effectively communicating results to stakeholders.

IST 691 | Deep Learning in Practice

The IST 691 Deep Learning in Practice course project is a prime example of how I have achieved the program goals outlined in the Syracuse University Master of Science in Applied Data Science program. This project showcases my ability to collect, store, and access data, create actionable insights, apply advanced machine learning techniques, use programming languages for data analysis, communicate insights effectively, and adhere to ethical considerations in data science practices.

The project involves developing an interactive Streamlit application that utilizes artificial neural networks to classify bird species based on their audio recordings. The dataset, consisting of recordings from five different bird species, was collected and preprocessed to ensure consistent sampling rates and audio channels. This demonstrates my proficiency in identifying and leveraging appropriate technologies for data collection and storage.

Figure 9

The Exploratory Data Analysis (EDA) section of the application illustrates the class imbalance of the dataset and allows users to explore the mel spectrograms associated with the audio recordings. Mel spectrograms are visual representations of the audio signals, highlighting the frequency content over time. By providing this feature, the application enables users to gain a deeper understanding of how the neural network utilizes these spectrograms to classify the bird species accurately. This showcases the overall ability to create actionable insights and communicate them effectively through interactive visualizations.

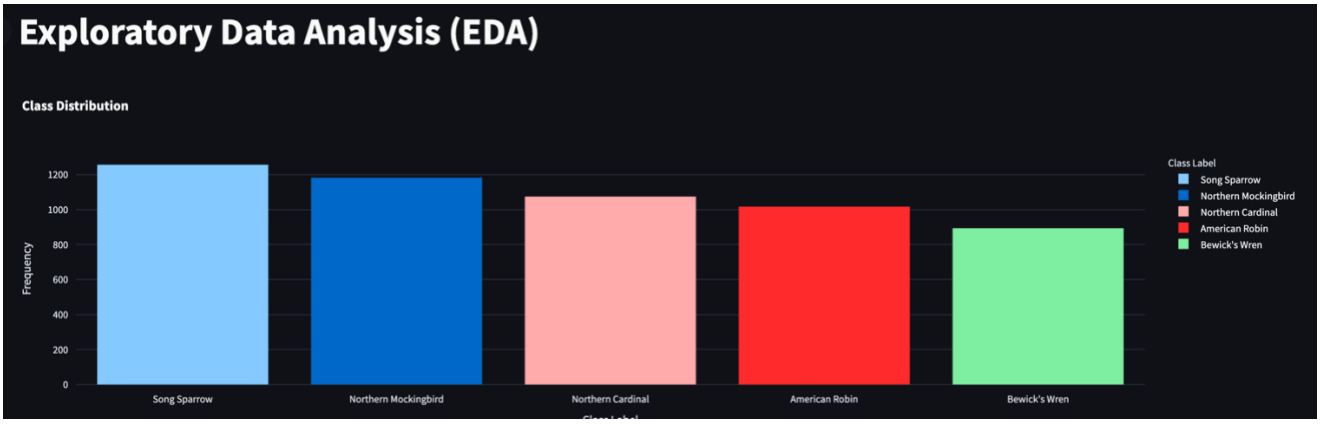

Figure 10.

The project focuses on evaluating the performance of a convolutional neural network (CNN) model architecture for audio classification. The application tests this model architecture under various configurations – including fixed learning rate, scheduled learning rate, early stopping, and stratified sampling of the class distributions. This demonstrates proficiency in applying advanced machine learning techniques, specifically deep learning, to solve real-world problems.

The Model Inference section of the application presents the performance metrics of the trained models, allowing users to compare the impact of different training strategies. The results highlight the effectiveness of techniques such as scheduled learning rate, early stopping, and stratified sampling in improving classification accuracy. By presenting these metrics in an interactive and user-friendly manner, I showcase my skills in effectively communicating the outcomes of deep learning models to a broad audience.

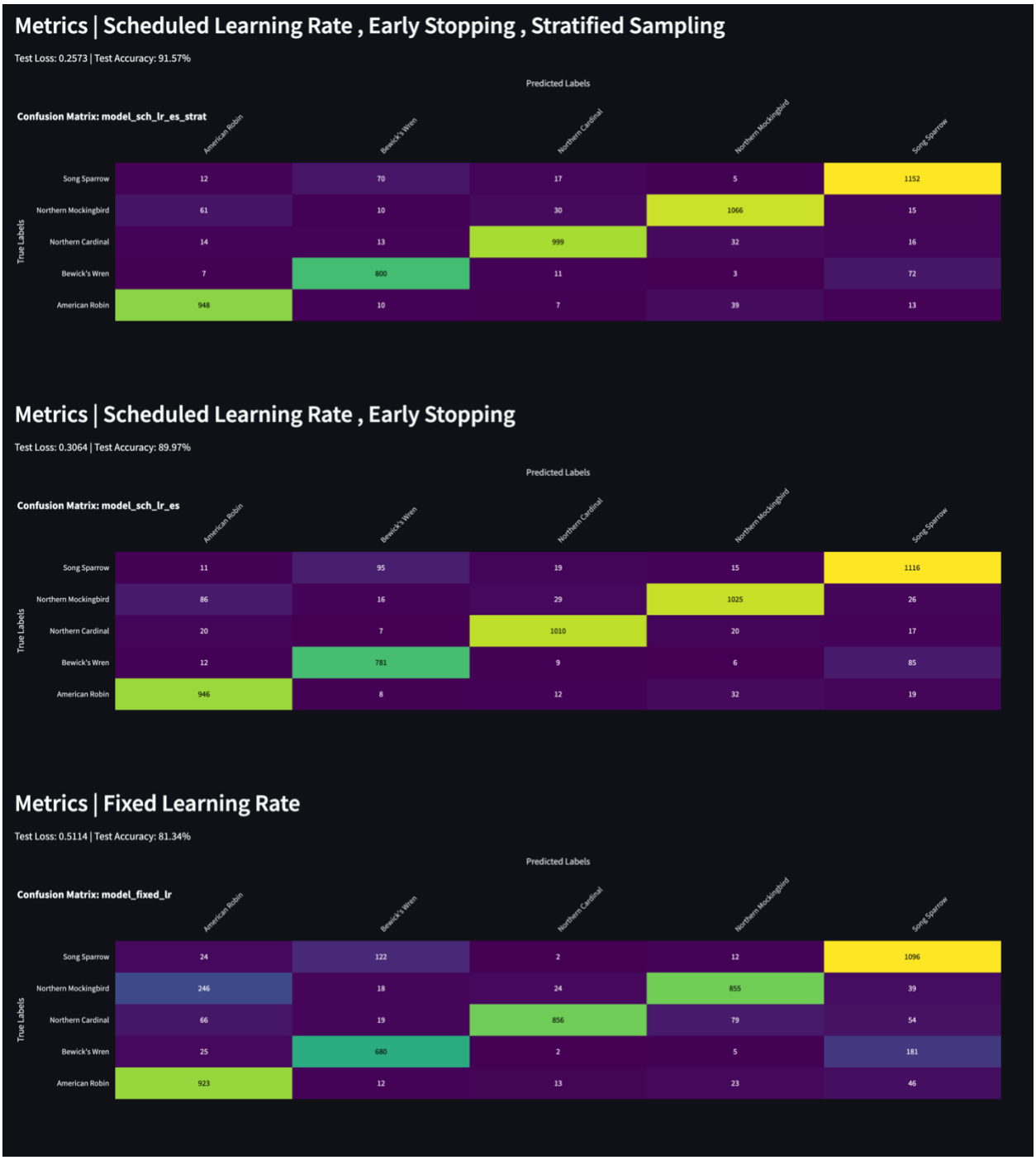

Figure 11.

Throughout the project, I have utilized PyTorch, a powerful deep learning framework, to develop and train the neural networks for bird audio classification. This demonstrates my ability to use programming languages and frameworks specifically designed for deep learning tasks, enabling me to build sophisticated models and achieve accurate results.

Furthermore, the Streamlit application itself highlights my proficiency in creating engaging and informative interfaces to present the results of deep learning projects. The application's features, such as the class distribution visualization, mel spectrogram exploration, and model performance comparison, provide users with a comprehensive understanding of the bird audio classification task and the underlying deep learning techniques employed.

An example of the mel spectrogram and audio recording can be seen below:

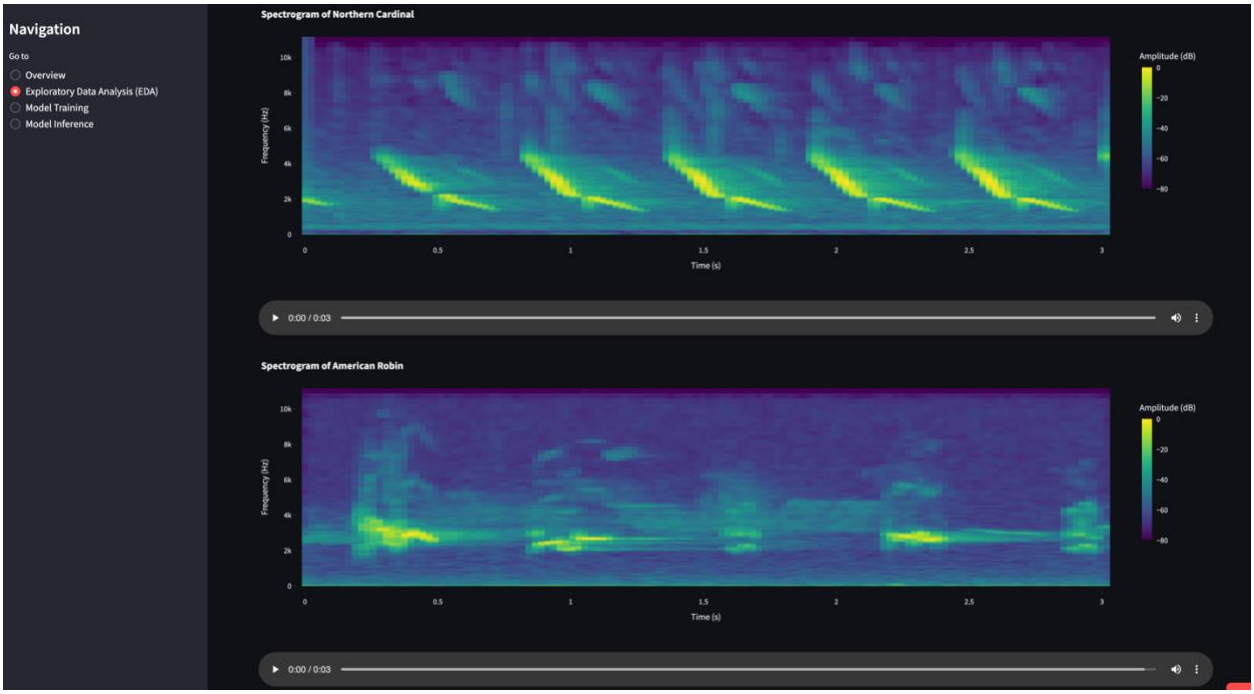

Figure 12.

In the above example, the audience can view the image representation of the bird song for the Northern Cardinal and American Robin's respective songs. An option to play a short, 3-second audio clip is available below the spectrogram.

By successfully completing the IST 691 Deep Learning in Practice course project, along with the IST 736 Text Mining and IST 718 Big Data Analytics projects, I have demonstrated my ability to achieve the key program goals and outcomes. These projects showcase my skills in data collection and preprocessing, applying advanced machine learning and deep

learning techniques, generating actionable insights, effectively communicating results, and adhering to ethical considerations in data science practices.

The IST 691 project, in particular, highlights my proficiency in leveraging deep learning techniques to solve real-world problems, utilizing programming languages and frameworks for data analysis, creating interactive visualizations to communicate insights, and developing user-friendly applications to make the results accessible to a wide audience. This project serves as tangible evidence of my readiness to tackle complex data science challenges and contribute to the field as a competent and skilled data science professional.

Conclusion

As I reflect on my journey through the Master of Science in Applied Data Science program at Syracuse University, I am filled with a sense of accomplishment and gratitude. This program has been instrumental in shaping my skills, knowledge, and perspective as a data science professional.

Through the coursework and projects, I have gained a deep understanding of the various facets of data science, from data collection and preprocessing to advanced machine learning techniques and big data analytics. The program's emphasis on practical application and real-world problem-solving has equipped me with the tools and mindset necessary to tackle complex data-driven challenges.

The projects I have completed, such as the IST 736 Text Mining project, the IST 718 Big Data Analytics project, and the IST 691 Deep Learning in Practice project, serve as testament to my ability to apply the knowledge and skills acquired throughout the program. These projects demonstrate my proficiency in leveraging cutting-edge technologies, applying appropriate methodologies, and generating actionable insights from data.

Moreover, the program has honed my ability to communicate complex technical concepts effectively to both technical and non-technical audiences. Through the development of interactive visualizations and user-friendly applications, I have learned to present insights in a clear and concise manner – enabling stakeholders to make data-driven decisions with confidence.

As I embark on the next chapter of my career, I am confident that the skills, knowledge, and experiences gained through the Master of Science in Applied Data Science program will serve as a solid foundation for my future endeavors. I am excited to apply my newfound skills to real-world problems, contribute to the advancement of the field, and make a positive impact on organizations and society as a whole.

I am grateful for the opportunity to have been a part of this transformative program and for the support and guidance provided by the faculty, staff, and my fellow classmates. I look forward to leveraging the network and resources available to me as an alumnus of Syracuse University and to continue my journey of lifelong learning and growth in the field of data science.

With the skills, knowledge, and mindset acquired through the Master of Science in Applied Data Science program, I am well-prepared to tackle the challenges and opportunities that lie ahead. I am excited to contribute to the data science community, drive innovation, and make a meaningful impact in the world through the power of data-driven insights.

Appendix Resume

Matthew L. Pergolski

240106 Del Rio Rd • Wausau, WI 54403
689-222-6968 • mlpergol@syr.edu • matthew.l.pergolski@lmco.com

Education

Syracuse University, Syracuse, NY
May 2024
M.S., Applied Data Science

- Grade: 4.0

University of Wisconsin-Eau Claire, Eau Claire, WI
December 2017
B.B.A., Operations/Supply Chain Management

- Graduated Cum Laude

Work Experience

Lockheed Martin, Remote **July 2022-Present**
Data Scientist Senior *(Full-Time)*

- Utilizes various Data Science tools to achieve business results: Domino Data Lab, Python, SQL, AWS
- Applies the following machine learning algorithms to business problems: Supervised regression/classification, natural language processing, deep learning, anomaly detection, reinforcement learning
- Demonstrates creativity and innovation in identifying valuable ways of incorporating machine learning into business workflows and processes
- Displays strong problem-solving skills, business acumen, and demonstrates excellent oral and written communication skills

- Collaborates with other scientists, engineers, consultants, and database administrators of all backgrounds and disciplines to bring analytical rigor and statistical methods to the challenges of predicting behaviors

Lockheed Martin, Remote **July 2022-September 2023**
Data Scientist *(Full-Time)*

- Utilized various Data Science tools to achieve business results: Domino Data Lab, Python, SQL, R, AWS
- Applied the following machine learning algorithms to business problems: Supervised regression/classification, natural language processing, deep learning, anomaly detection, reinforcement learning
- Demonstrated creativity and innovation in identifying valuable ways of incorporating machine learning into business workflows and processes
- Displayed strong problem-solving skills, business acumen, and demonstrates excellent oral and written communication skills
- Collaborates with other scientists, engineers, consultants, and database administrators of all backgrounds and disciplines to bring analytical rigor and statistical methods to the challenges of predicting behaviors

Lockheed Martin, Orlando, FL **November 2019-July 2022**
Manufacturing Planner *(Full-Time)*

- Performed the bill of material (BOM) planning activities associated with manufacturing resource planning (MRP) transactions for the following programs: LRASM, JAGM, Javelin
- Executed Capacity Requirements Planning (CRP) on a quarterly basis; collaborated with Industrial Engineering team to analyze and communicate results
- Served as the facilitator for the Orlando-Program Lead Enterprise (0PLE) Performance Management Team (PMT); set and determined meeting agendas, consulted with metric owners, and provided support to the PMT lead
- Amassed $200,000+ in cost-avoidance savings for Orlando-Bill of Material (0BOM) PMT through optimizing reporting structures in SAP-ERP system

Lockheed Martin, Orlando, FL **January 2018-November 2019**
Manufacturing Planner Associate *(Full-Time)*

- Performed the buy, make, and bill of material planning activities associated with manufacturing resource planning transactions for the following programs: Special Programs, IRST, MLD
- Maintained production systems integrity to drive inventory optimization to minimize cost
- Sought to expedite and optimize material flow through the supply chain to meet customer requirements
- Led Performance Management Team (PMT) Savings team in generating concrete ideas for process improvement; derived a total of ~$180,000 in total cost-avoidance savings
- Contributed to planner PMT group as a dedicated and detailed owner of the following metric: Tier 2 Scheduled Completions; collaborated with PP&C leads of 14 programs to convey monthly results to management

Lockheed Martin, Ocala, FL **May 2017-August 2017**
Manufacturing Planner Intern *(Full-Time)*

- Led weekly PMT meetings with both salary and hourly employees; recapped weekly team metric goals and provided team building activities for Surface-Mount Technology (SMT) group

- Collaborated with Lockheed Martin supplier officials from Victory Supply to improve processes relating to consumable inventory; implemented inventory referencing system resulting in an estimated $120,000 in total cost avoidances

Additional Information

Software

- Application Software: R, Python, SQL, SAP HANA, Tableau, Git, JIRA
- Production Software: SAP, Windchill (dPLM), APECS
- Other: Microsoft Office Suite (i.e., Excel, Power Query, Word, PowerPoint, Project, Visio), draw.io, Weka

Other Qualifications

- Active U.S. Government Security Clearance
- Coursera Completion Certificates (i.e., Python Basics, R Programming, Data Visualization with dplyr and ggplot2 in R, Data Manipulation with dplyr in R)
- LinkedIn Learning Completion Certificates (i.e., Git and Gitlab, Tableau, Microsoft Project, JIRA, R Programming)
- Lean Six Sigma Green Belt Certified
- Excellent written communication, time management skills