IT SKILL STANDARDS 2020 AND BEYOND

"Data Analytics and Predictive Modeling"

Job Cluster

Acknowledgements

The development and publication of these skill standards has been a joint and collaborative effort between business and industry representatives and the education community. We are grateful to the industry personnel who participated in the development and validation process. Industry subject matter experts, technical executives, supervisors and technicians donated their time and effort to assure the relevancy of the standards 12 to 36 months into the future.

We gratefully acknowledge funding from the National Science Foundation and the leadership by the team on the IT Skill Standards 2020 and Beyond grant, based at Collin College.

Our leaders are strategically divided into Central, Western, and Eastern teams.

Central

Dr. Ann Beheler, Principal Investigator

Christina Titus, Program Director

Deborah Roberts, Co-Principal Investigator

Helen Sullivan, Senior Staff

West Coast

Terryll Bailey, Co-Principal Investigator

Dr. Suzanne Ames, Co-Principal Investigator

East Coast

Peter Maritato, Co-Principal Investigator

Gordon Snyder, Senior Staff

This material is based upon work supported by the National Science Foundation under Grant No. 1838535. Any opinions, findings and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

itskillstandards.org

Data Analytics and Predictive Modeling

The definition for Data Analytics and Predictive Modeling as developed by approximately 100 Thought Leaders (mostly Chief Technology Officers and Chief Information Officers) through three meetings and follow-up surveys to gain consensus is:

> Data Analytics and Predictive Modeling includes inspecting, cleansing, transforming, and modeling data with the goal of discovering useful information, informing conclusions, and supporting decision-making. Business intelligence (BI) specifically focuses on extracting business information for use by decision makers. Common functions of business intelligence include reporting, data mining, process mining, benchmarking, and text mining. This definition was adapted from Wikipedia with input from IT Thought Leaders.

This packet includes…

Job skills as developed by subject matter experts (SMEs) via multiple synchronous meetings (Page 3).

The tasks, knowledge, skills and abilities (KSAs) were developed with a focus 12 to 36 months in the future for an entry-level employee working in that specific cluster.

More specific definitions can be found within the KSA list.

The average was calculated from the subject matter expert votes.

- A vote of "4" indicated the item must be covered in the curriculum.
- A vote of "3" indicated the item should be covered in the curriculum.
- A vote of "2" indicated that it would be nice for the item to be covered in the curriculum.
- A vote of "1" indicated the item should not be covered in the curriculum.

Employability Skills as developed by SMEs via multiple synchronous meetings (Page 7).

Employability competencies are essential for every IT job and are based on what the work requires. SMEs were offered three clearly-defined “levels of proficiency” for each employability skill. The proficiency scale is defined as Level 1 – basic; Level 2- intermediate; and Level 3 - advanced. The levels are cumulative, so a “Level 3” assumes the employee can perform all characteristics of “Level 1” and “Level 2.”

For each employability skill, SMEs selected the competency levels that best aligned with what would be expected from an entry-level worker for the job cluster in question.

Key Performance Indicators (KPIs) as developed by SMEs (Page 8).

Key Performance Indicators answer the question, “How do we know when a task is performed well?”

A search was performed to locate validated/verified KPIs for technician level work in IT fields. Sources included the Texas Skill Standards System, National Skill Standards Board, National Institute of Standards and Technology and other sources. The identified KPIs were then cross-referenced to the tasks for the

ITSS 2020 job clusters. They were reviewed and revised by a group of the same subject matter experts who developed the tasks and KSAs for the cluster in a structured, facilitated verification session.

Student Learning Outcomes (SLOs) as identified by educators attending the KSA meetings (Page 10).

The SLOs are for use in the creation of curriculum to help define what the students will know and be able to demonstrate. Each of these SLOs can be observed, measured, and demonstrated.

Data Analytics and Predictive Modeling Tasks and KSAs		
Task SPECIFIC THINGS an entry level person would BE EXPECTED TO PERFORM on the job WITH LITTLE SUPERVISION.		**AVG**
Business Problem (Question) Framing		
T-1	Assist in obtaining or receiving problem statement and usability requirements.	3.6
T-2	Assist in identifying stakeholders.	3.0
T-3	Assist in determining if the problem is amenable to an analytics solution.	3.3
T-4	Assist in refining the problem statement and delineate.	3.2
T-5	Assist in defining an initial set of business benefits.	3.2
T-6	Assist in obtaining stakeholder agreement on the problem.	3.1
Analytics Problem Framing		
T-7	Assist in reformulating the problem statement as an analytics problem.	3.4
T-8	Assist in developing a proposed set of drivers and relationships to outputs.	3.1
T-9	Assist in stating the set of assumptions related to the problem.	3.3
T-10	Assist in defining key metrics of success.	3.4
T-11	Assist with collecting metrics and trending data.	3.2
T-12	Assist in obtaining stakeholder agreement on analytical approach.	2.9
Data		
T-13	Assist with identifying and prioritizing data needs and sources.	3.3
T-14	Assist with assessing the validity of source data and subsequent findings.	3.0
T-15	Assist in acquiring data.	3.3
T-16	Assist in harmonizing, rescaling, cleaning, and sharing data.	3.5
T-17	Assist with identifying relationships in the data.	3.3
T-18	Assist with documenting and reporting findings (e.g., insights, results, business performance).	3.5
T-19	Assist with refining the business and analytics problem statements.	3.2
Methodology (Approach) Selection		
T-20	Assist with identifying available problem solving approaches (methods).	3.3
T-21	Assist in conferring with systems analysts, engineers, programmers, and others to design application.	2.8
T-22	Assist in using software tools.	2.8
T-23	Assist in reading, interpreting, writing, modifying, and executing simple scripts (e.g., Perl, VBScript) on Windows and UNIX systems (e.g., those that perform tasks such as: parsing large data files, automating manual tasks, and fetching/processing remote data).	3.2
T-24	Assist in utilizing different programming languages to write code, open files, read files, and write output to different files.	3.2
T-25	Assist in utilizing open source language such as R and apply quantitative techniques (e.g., descriptive and inferential statistics, sampling, experimental design, parametric and non-parametric tests of difference, ordinary least squares regression, general line).	3.2
T-26	Assist with developing and implementing data mining and data programs.	2.7
T-27	Assist with testing approaches (methods).	3.4
T-28	Assist in conducting hypothesis testing using statistical processes.	3.1
T-29	Assist with providing analyses and support for effectiveness assessment.	2.8
T-30	Assist with selecting approaches (methods).	3.3
Model Building		
T-31	Assist with identifying model structures.	3.1
T-32	Assist in running and evaluating the models.	3.4
T-33	Assist with tuning models and data.	3.2
T-34	Assist with integrating the models.	2.9
T-35	Assist with documenting and communicating findings (including assumptions, limitations and constraints).	3.7
T-36	Assist with performing internal business verification and validation of the model.	3.0
T-37	Assist with publishing validation and verification report.	2.9
T-38	Assist in developing recommendations to the supervisor based on data analysis and findings.	3.1
Deployment		

T-39	Assist with deploying application codes and analytical models using CI/CD tools and techniques and provides support for deployed data applications and analytical models.	2.6
T-40	Assist with performing business validation of the model.	3.1
T-41	Assist with presenting technical information to technical and nontechnical audiences.	3.3
T-42	Assist with presenting data in creative formats.	2.8
T-43	Assist with delivering reports with findings.	3.5
T-44	Assist with creating model, usability, and system requirements for production.	2.8
T-45	Assist in supporting deployment.	2.8
Model Lifecycle Management		
T-46	Assist with documenting initial structure.	3.0
T-47	Assist in tracking model quality.	3.3
T-48	Assist with providing input and assist in post-action effectiveness assessments.	2.7
T-49	Assist in the identification of information collection shortfalls.	2.8
T-50	Assist with recalibrating and maintaining the model.	3.0
T-51	Assist with evaluating the business benefit of the model over time.	2.9
T-52	Assist with developing strategic insights from large data sets.	2.9
Knowledge Knowledge focuses on the understanding of concepts. It is theoretical. An individual may have an understanding of a topic or tool or some textbook knowledge of it but have no experience applying it. For example, someone might have read hundreds of articles on health and nutrition, many of them in scientific journals, but that doesn't make that person qualified to dispense advice on nutrition.		
K-1	Knowledge of risk management processes (e.g., methods for assessing and mitigating risk).	2.7
K-2	Knowledge of computer algorithms.	3.3
K-3	Knowledge of computer programming principles.	3.2
K-4	Knowledge of data administration and data standardization policies.	2.9
K-5	Knowledge of data mining and data management principles.	3.0
K-6	Knowledge of database management systems, query languages, table relationships, and views.	3.1
K-7	Knowledge of mathematics (e.g., logarithms, trigonometry, linear algebra, calculus, statistics, and operational analysis).	3.3
K-8	Knowledge of programming language structures and logic.	3.2
K-9	Knowledge of query languages such as SQL (structured query language).	3.5
K-10	Knowledge of sources, characteristics, and uses of the organization’s data assets.	2.9
K-11	Knowledge of the various technologies for organizing and managing information (e.g., databases, bookmarking engines).	2.8
K-12	Knowledge of command-line tools (e.g., mkdir, mv, ls, passwd, grep).	2.7
K-13	Knowledge of interpreted and compiled computer languages.	2.7
K-14	Knowledge of how to utilize Hadoop, Java, Python, SQL, Hive, and Pig to explore data.	3.2
K-15	Knowledge of machine learning theory and principles.	3.3
K-16	Knowledge of data classification standards and methodologies based on sensitivity and other risk factors.	2.7
K-17	Knowledge of Personally Identifiable Information (PII) data security standards.	2.9
K-18	Knowledge of the principal methods, procedures, and techniques of gathering information and producing, reporting, and sharing information.	3.1
K-19	Knowledge of data mining techniques.	3.4
K-20	Knowledge of database theory.	2.8
K-21	Knowledge of how to extract, analyze, and use metadata.	3.0
K-22	Knowledge of ETL techniques, Hadoop, Data analytics, Big data is an advantage.	2.9
K-23	Knowledge of a variety of machine learning techniques (clustering, decision tree learning, artificial neural networks, etc.) and their real-world advantages/drawbacks.	3.2
K-24	Knowledge of advanced statistical techniques and concepts (regression, properties of distributions, statistical tests and proper usage, etc.) and experience with applications.	3.2
K-25	Knowledge of the underlying theory and concepts of Relational Databases (e.g., Microsoft SQL Server, Oracle, Teradata MySQL).	2.8
K-26	Knowledge of Decision Science Game theory.	2.9

K-27	Knowledge of the use of simulation.	3.1
K-28	Knowledge of optimization.	3.3
K-29	Knowledge of data analysis concepts.	3.6
K-30	Knowledge of how to identify and document potential ethical concerns for application of model outputs.	3.1
Skills The capabilities or proficiencies developed through training or hands-on experience. Skills are the practical application of theoretical knowledge. Someone can take a course to gain knowledge of concepts without developing the skills to apply those concepts. Development of skills requires hands-on application of the concepts.		
S-1	Skill in conducting queries and developing algorithms to analyze data structures.	3.5
S-2	Skill in creating and utilizing mathematical or statistical models.	3.3
S-3	Skill in data mining techniques (e.g., searching file systems) and analysis.	3.3
S-4	Skill in using and contributing content to data dictionaries.	2.7
S-5	Skill in developing data models.	3.0
S-6	Skill in generating queries and reports.	3.5
S-7	Skill in writing code in a currently supported programming language (e.g., Python).	2.9
S-8	Skill in data pre-processing (e.g., imputation, dimensionality reduction, normalization, transformation, extraction, filtering, smoothing).	3.0
S-9	Skill in identifying patterns or relationships.	3.1
S-10	Skill in performing sentiment analysis.	3.3
S-11	Skill in Regression Analysis (e.g., Hierarchical Stepwise, Generalized Linear Model, Ordinary Least Squares, Tree-Based Methods, Logistic).	3.1
S-12	Skill in supporting transformation analytics to invoke a business shift.	2.9
S-13	Skill in using basic descriptive statistics and techniques (e.g., normality, model distribution, scatter plots).	3.4
S-14	Skill in using data analysis tools (e.g., Excel, Python).	3.3
S-15	Skill in using data mapping tools.	2.9
S-16	Skill in using outlier identification and removal techniques.	3.3
S-17	Skill in writing scripts using R, Python, PIG, HIVE, SQL, etc.	3.5
S-18	Skill to identify sources, characteristics, and uses of the organization's data assets.	2.8
S-19	Skill in conducting information searches.	3.0
S-20	Skill in developing or recommending analytic approaches or solutions to problems and situations for which information is incomplete or for which no precedent exists.	2.8
S-21	Skill in evaluating information for reliability, validity, and relevance.	2.9
S-22	Skill in preparing and presenting briefings.	3.1
S-23	Skill in tailoring analysis to the necessary levels (e.g., classification and organizational).	2.9
S-24	Skill in using multiple search engines (e.g., Google, Yahoo, LexisNexis, DataStar) and tools in conducting open-source searches.	2.8
S-25	Skill in utilizing feedback to improve processes, products, and services.	3.1
S-26	Skill in performing data analysis including applying statistics.	3.6
S-27	Skill in using statistical computer languages (R, Python, etc.) to manipulate data and draw insights from large data sets.	3.5
S-28	Skill in Visualization using R, Python, or other languages and frameworks.	3.4
S-29	Skill in problem-solving skills and critical thinking ability.	3.6
S-30	Skill in collaboration and communication skills within and across teams.	3.6
S-31	Skill in analytics problem framing (e.g., define geometric sets).	3.5
Abilities Abilities have historically been used to describe the innate traits or talents that a person brings to a task or situation. Many people can learn to negotiate competently by acquiring knowledge about it and practicing the skills it requires. A few are brilliant negotiators because they have the innate ability to persuade. In reality, abilities may be included under skills or may be separated out.		
A-1	Ability to dissect a problem and examine the interrelationships between data that may appear unrelated.	3.2
A-2	Ability to identify basic common coding flaws at a high level.	3.0
A-3	Ability to use data visualization tools (e.g., Flare, HighCharts, AmCharts, D3.js, Processing, Google Visualization API, Tableau, Raphael.js).	3.3

A-4	Ability to source data used in information, assessment, and/or planning products.	2.7
A-5	Ability to communicate complex information, concepts, or ideas in a confident and well-organized manner through verbal, written, and/or visual means.	3.5
A-6	Ability to develop or recommend analytic approaches or solutions to problems and situations for which information is incomplete or for which no precedent exists.	3.0
A-7	Ability to evaluate, analyze, and synthesize large quantities of data (which may be fragmented and contradictory) into quality, fused targeting/information products.	2.9
A-8	Ability to clearly articulate information requirements into well-formulated research questions and data tracking variables for inquiry tracking purposes.	2.9
A-9	Ability to effectively collaborate via virtual teams.	3.3
A-10	Ability to evaluate information for reliability, validity, and relevance.	3.3
A-11	Ability to exercise strong ethical judgment when policies are not well-defined.	2.8
A-12	Ability to focus research efforts to meet the customer's decision-making needs.	3.3
A-13	Ability to adapt to a dynamic environment.	3.2
A-14	Ability to function in a collaborative environment, seeking continuous consultation with other analysts and experts—both internal and external to the organization—to leverage analytical and technical expertise.	3.6
A-15	Ability to identify information gaps.	3.1
A-16	Ability to recognize and mitigate cognitive biases which may affect analysis.	3.1
A-17	Ability to recognize and mitigate deception in reporting and analysis.	2.9
A-18	Ability to think critically.	3.8
A-19	Ability to understand objectives and effects.	3.3
A-20	Ability to utilize multiple information sources across all information disciplines.	3.2
A-21	Ability to effectively communicate ideas to team members with varying levels of technical expertise.	3.7
A-22	Ability to understand a business problem.	3.7
A-23	Ability to understand and use the databases and tools to run queries to solve the business problem.	3.7
A-24	Ability to identify patterns.	3.3

Data Analytics and Predictive Modeling Employability Skills

Workplace Professionalism and Work Ethics	Level 1 - Employee learns expectations of workplace environment (professional behavior and ethics) and adheres to practices with some guidance. Level 2 - Employee exhibits sound professionalism, judgment, and integrity and accepts responsibility for own behavior. Employee exhibits these qualities without guidance but occasionally refers to policies as needed.
Written Communication	Level 1 - Employee understands written instructions and executes tasks with guidance and feedback from supervisor. Employee clearly communicates concepts in writing. Level 2 - Employee comprehends and executes written instructions with minimal guidance. Employee composes well-organized written documents.
Oral Communication	Level 1 - Employee understands oral instructions and executes tasks with guidance and feedback from supervisor. Employee communicates concepts orally while clarifying for meaning. Employee develops listening skills. Level 2 - Employee comprehends and executes oral instructions with minimal guidance and exhibits good listening skills. Employee clarifies for meaning without needing prompting from supervisor.
Teamwork	Level 1 - With guidance and feedback from supervisor, employee obeys team rules and understands team member roles. Employee actively participates in team activities, volunteers for special tasks, and establishes rapport with co-workers.
Problem Solving & Critical Thinking	Level 1 - Employee identifies the problem and relevant facts and principles with guidance and feedback from supervisor. Employee summarizes existing ideas and demonstrates creative thinking process while problem solving.
Organization and Planning	Level 1 - Employee prepares schedule for self, monitors and adjusts task sequence, and analyzes work assignments with guidance from supervisor. Level 2 - Employee manages timelines and recommends timeline adjustments. Employee escalates timeline-impacting issues as appropriate.
Adaptability and Flexibility	Level 1 - With guidance and feedback from supervisor, employee is able to adjust ways of doing work based on changing dynamics. Working under pressure is difficult, but employee makes it through the project with guidance and oversight.
Initiative	Level 1 - Employee finishes a step in a project and waits for direction before going on to the next step. Level 2 - Employee finishes multiple steps in a project and appropriately begins working on the next step without being asked.
Accuracy	Level 1 - Employee makes mistakes routinely but is committed to learning to adjust work habits to prevent them in the future. Level 2 - Employee occasionally makes mistakes but quickly makes adjustments to work habits to avoid making the same mistake twice.
Cultural Competence	Level 1 - Employee is inexperienced with working with diverse teams. With support and guidance and getting to know team members, employee develops working relationships. Level 2 - Employee is committed to working with diverse teams but struggles when differences arise. Employee identifies those challenges and works with colleagues to find ways to work effectively.
Self and Career Development	Level 1 - Employee requires feedback and direction from supervisor regarding improvement needed in professional and technical skills. Employee follows through with skills development with monitoring by supervisor.

Data Analytics and Predictive Modeling Key Performance Indicators

For the entry-level employee, all tasks are typically done under supervision for much of the first year and then with some independence with verification after the employee has more experience. All tasks are done according to company guidelines.

	Task	Key Performance Indicators
Business Problem (Question) Framing		
T-1	Assist in obtaining or receiving problem statement and usability requirements.	Appropriate stakeholders are identified in a timely manner. Problem statement and usability requirements are obtained in a timely manner and properly documented. Determination of the applicability of an analytics solution is accurate. Business and analytics problem statements are clear, and are continuously refined. Business benefits are correctly identified and clearly stated.
T-2	Assist in identifying stakeholders.	
T-3	Assist in determining if the problem is amenable to an analytics solution.	
T-4	Assist in refining the problem statement and delineate.	
T-5	Assist in defining an initial set of business benefits.	
T-6	Assist in obtaining stakeholder agreement on the problem.	
Analytics Problem Framing		
T-7	Assist in reformulating the problem statement as an analytics problem.	The alternatives to the analytics problem statement are documented and ranked according to best match with current problem and rationale for choices clearly stated. Assumptions related to the problem are stated clearly and concisely. Criteria for success are clearly identified. Agreement of stakeholders is obtained regarding business and analytics problem statements and analytic approach.
T-8	Assist in developing a proposed set of drivers and relationships to outputs.	
T-9	Assist in stating the set of assumptions related to the problem.	
T-10	Assist in defining key metrics of success.	
T-11	Assist with collecting metrics and trending data.	
T-12	Assist in obtaining stakeholder agreement on analytical approach.	
Data		
T-13	Assist with identifying and prioritizing data needs and sources.	Sources and methods for acquiring data are efficient and information is accurate and complete. Data is secured from reliable and respected sources. Data is correctly harmonized, rescaled, and cleaned and relationships in the data are correctly identified. Findings are documented in accordance with company procedures and communicated in a clear and timely manner. Data definitions are fully developed and agreed upon in accordance with company procedures.
T-14	Assist with assessing the validity of source data and subsequent findings.	
T-15	Assist in acquiring data.	
T-16	Assist in harmonizing, rescaling, cleaning and sharing data.	
T-17	Assist with identifying relationships in the data.	
T-18	Assist with documenting and reporting findings (e.g., insights, results, business performance).	
T-19	Assist with refining the business and analytics problem statements.	
Methodology (Approach) Selection		
T-20	Assist with identifying available problem solving approaches (methods).	Sources and methods for acquiring data are efficient and information is accurate and complete. The alternatives to the methodology are documented and ranked. Data is secured from reliable and respected sources. Findings are documented in accordance with company procedures and communicated in a clear and timely manner. Data definitions are fully developed and agreed upon in accordance with company procedures. Problem solving approaches and methods are affordable and relevant. Analysis processes and conclusions are clearly and concisely documented. Effective software tools and problem-solving methods are used. Scripts are complete, relevant and congruent. Appropriate testing methodology is identified and planned and scope of testing is clearly identified. Algorithms, programming principles, statistical processes are used correctly.
T-21	Assist in conferring with systems analysts, engineers, programmers, and others to design application.	
T-22	Assist in using software tools.	
T-23	Assist in reading, interpreting, writing, modifying, and executing simple scripts (e.g., Perl, VBScript) on Windows and UNIX systems (e.g., those that perform tasks such as: parsing large data files, automating manual tasks, and fetching/processing remote data).	
T-24	Assist in utilizing different programming languages to write code, open files, read files, and write output to different files.	
T-25	Assist in utilizing open source language such as R and apply quantitative techniques (e.g., descriptive and inferential statistics, sampling, experimental design, parametric and non-parametric tests of difference, ordinary least squares regression, general line).	
T-26	Assist with developing and implementing data mining and data programs.	
T-27	Assist with testing approaches (methods).	
T-28	Assist in conducting hypothesis testing using statistical processes.	
T-29	Assist with providing analyses and support for effectiveness assessment.	
T-30	Assist with selecting approaches (methods).	
Model Building		
T-31	Assist with identifying model structures.	Models are evaluated, tuned and integrated using the proper procedures. Data model is laid out clearly. Performance criteria for the data model have verifiable assumptions. Scope and purpose of model are defined. Code is developed using efficient software design processes. Reusable components are employed whenever possible. Code is well documented so that it can be understood by others. Tests accurately assess the functions the module is designed to perform. Ethics reviews are routinely accomplished.
T-32	Assist in running and evaluating the models.	
T-33	Assist with tuning models and data.	
T-34	Assist with integrating the models.	
T-35	Assist with documenting and communicating findings (including assumptions, limitations and constraints).	
T-36	Assist with performing internal business verification and validation of the model.	
T-37	Assist with publishing validation and verification report.	
T-38	Assist in developing recommendations to the supervisor based on data analysis and findings.	

Deployment		
T-39	Assist with deploying application codes and analytical models using CI/CD tools and techniques and provides support for deployed data applications and analytical models.	Application codes and analytical models are deployed according to plan. Business validation of the model is performed correctly. Presentations are well-organized, utilize creative formats and meet the needs of technical and non-technical audiences. Enterprise goals are taken into account when drawing conclusions from data analysis and making recommendations to supervisor. Model, usability and system requirements for production are developed in accordance with company procedures. Requirements are properly interpreted and evaluated, and conflicting requirements are identified and resolved.
T-40	Assist with performing business validation of the model.	
T-41	Assist with presenting technical information to technical and nontechnical audiences.	
T-42	Assist with presenting data in creative formats.	
T-43	Assist with delivering reports with findings.	
T-44	Assist with creating model, usability, and system requirements for production.	
T-45	Assist in supporting deployment.	
Model Lifecycle Management		
T-46	Assist with documenting initial structure.	Initial structure of the model is documented in accordance with company standards and in a timely manner. Tracking of model quality and model recalibration and maintenance. Effectiveness testing is based on specification criteria. Recommendations are fed back into the modeling process. Computer data administration, data standardization, data mining and data management are conducted in accordance with industry and company procedures and standards.
T-47	Assist in tracking model quality.	
T-48	Assist with providing input and assist in post-action effectiveness assessments.	
T-49	Assist in the identification of information collection shortfalls.	
T-50	Assist with recalibrating and maintaining the model.	
T-51	Assist with evaluating the business benefit of the model over time.	
T-52	Assist with developing strategic insights from large data sets.	

Data Analytics and Predictive Modeling Student Learning Outcomes

Knowledge		Student Learning Outcomes
K-1	Knowledge of risk management processes (e.g., methods for assessing and mitigating risk).	Explain information security fundamentals. Demonstrate an understanding of the importance of ethics and privacy with data. Describe the functions of database recovery, security and administration, and basic data warehousing concepts.
K-16	Knowledge of data classification standards and methodologies based on sensitivity and other risk factors.	
K-17	Knowledge of Personally Identifiable Information (PII) data security standards.	
K-30	Knowledge of how to identify and document potential ethical concerns for application of model outputs.	
K-4	Knowledge of data administration and data standardization policies.	Describe the principles, techniques, and business policies for collecting, organizing, managing, analyzing, and reporting information. Describe the process of data science analytics from data acquisition to recommendations based on data. Describe different methods and tools for data collection and their impact on analysis of data. Identify the concepts of the relational model, normalization, dependencies, integrity, and constraints.
K-11	Knowledge of the various technologies for organizing and managing information (e.g., databases, bookmarking engines).	
K-18	Knowledge of the principal methods, procedures, and techniques of gathering information and producing, reporting, and sharing information.	
K-5	Knowledge of data mining and data management principles.	
K-19	Knowledge of data mining techniques.	
K-26	Knowledge of Decision Science Game theory.	
K-28	Knowledge of optimization.	
K-29	Knowledge of data analysis concepts.	
K-2	Knowledge of computer algorithms.	Apply the basics of programming principles. Demonstrate problem solving skills by developing and implementing algorithms to solve problems. Explain and apply the basic concepts of simulation-based methods.
K-3	Knowledge of computer programming principles.	
K-8	Knowledge of programming language structures and logic.	
K-12	Knowledge of command-line tools (e.g., mkdir, mv, ls, passwd, grep).	
K-13	Knowledge of interpreted and compiled computer languages.	
K-27	Knowledge of the use of simulation.	
K-6	Knowledge of database management systems, query languages, table relationships, and views.	Identify different database systems.
K-9	Knowledge of query languages such as SQL (structured query language).	
K-7	Knowledge of mathematics (e.g., logarithms, trigonometry, linear algebra, calculus, statistics, and operational analysis).	Select appropriate mathematical and statistical tools used for data analytics.
K-24	Knowledge of advanced statistical techniques and concepts (regression, properties of distributions, statistical tests and proper usage, etc.) and experience with applications.	
K-10	Knowledge of sources, characteristics, and uses of the organization's data assets.	Describe the data acquisition process. Explain data warehousing architectures, processes, and operations.
K-21	Knowledge of how to extract, analyze, and use metadata.	
K-14	Knowledge of how to utilize Hadoop, Java, Python, SQL, Hive, and Pig to explore data.	Describe tools and techniques to store and process data.
K-22	Knowledge of ETL techniques, Hadoop, Data analytics, Big data is an advantage.	
K-15	Knowledge of machine learning theory and principles.	Explain machine learning principles and techniques.
K-23	Knowledge of a variety of machine learning techniques (clustering, decision tree learning, artificial neural networks, etc.) and their real-world advantages/drawbacks.	
K-20	Knowledge of database theory.	Discuss database tools and techniques.
K-25	Knowledge of the underlying theory and concepts of Relational Databases (e.g., Microsoft SQL Server, Oracle, Teradata MySQL).	
Skills		**Student Learning Outcomes**
S-1	Skill in conducting queries and developing algorithms to analyze data structures.	Perform queries and develop reports.
S-6	Skill in generating queries and reports.	
S-3	Skill in data mining techniques (e.g., searching file systems) and analysis.	Create data models and use data mining techniques, models and tools. Cleanse and prepare data for analysis.
S-4	Skill in using and contributing content to data dictionaries.	
S-5	Skill in developing data models.	
S-15	Skill in using data mapping tools.	
S-16	Skill in using outlier identification and removal techniques.	
S-8	Skill in data pre-processing (e.g., imputation, dimensionality reduction, normalization, transformation, extraction, filtering, smoothing).	
S-9	Skill in identifying patterns or relationships.	Develop or discover analytical patterns from data models.
S-10	Skill in performing sentiment analysis.	

S-11	Skill in Regression Analysis (e.g., Hierarchical Stepwise, Generalized Linear Model, Ordinary Least Squares, Tree-Based Methods, Logistic).	Design and develop analytical solutions using appropriate mathematical and statistical models and tools. Apply data analytic tools to data in order to predict outcomes and classify data. Develop a software statistical modeling project and present the solution.
S-12	Skill in supporting transformation analytics to invoke a business shift.	
S-13	Skill in using basic descriptive statistics and techniques (e.g., normality, model distribution, scatter plots).	
S-26	Skill in performing data analysis including applying statistics.	
S-2	Skill in creating and utilizing mathematical or statistical models.	
S-31	Skill in analytics problem framing (e.g., define geometric sets).	
S-23	Skill in tailoring analysis to the necessary levels (e.g., classification and organizational).	
S-14	Skill in using data analysis tools (e.g., Excel, Python).	Demonstrate coding and scripting techniques using data analytics programming languages. Use appropriate programming language, data structures, and concepts to solve data science problems. Build arrays, data frames, dictionaries and perform basic calculations using programming languages. Develop a software statistical modeling project and present the solution. Use data visualization tools to analyze data and produce reports.
S-7	Skill in writing code in a currently supported programming language (e.g., Python).	
S-17	Skill in writing scripts using R, Python, PIG, HIVE, SQL, etc.	
S-27	Skill in using statistical computer languages (R, Python, etc.) to manipulate data and draw insights from large data sets.	
S-28	Skill in Visualization using R, Python, or other languages and frameworks.	
S-18	Skill to identify sources, characteristics, and uses of the organization's data assets.	Identify and interpret the data relevance, reliability, and validity from multiple sources. Describe the business intelligence methodology and concepts and relate them to decision support.
S-24	Skill in using multiple search engines (e.g., Google, Yahoo, LexisNexis, DataStar) and tools in conducting open-source searches.	
S-19	Skill in conducting information searches.	
S-20	Skill in developing or recommending analytic approaches or solutions to problems and situations for which information is incomplete or for which no precedent exists.	
S-21	Skill in evaluating information for reliability, validity, and relevance.	
S-25	Skill in utilizing feedback to improve processes, products, and services.	Demonstrate effective collaboration and communication skills to improve team productivity. Derive problem specifications from problem statements.
S-29	Skill in problem-solving skills and critical thinking ability.	
S-30	Skill in collaboration and communication skills within and across teams.	
S-22	Skill in preparing and presenting briefings.	
Abilities		**Student Learning Outcomes**
A-1	Ability to dissect a problem and examine the interrelationships between data that may appear unrelated.	Proficient in solving business problems by identifying data gaps and synthesizing data to deliver quality output. Utilize SQL and QBE commands to define, query and manipulate a relational database. Apply databases to actual situations and business problems.
A-4	Ability to source data used in information, assessment, and/or planning products.	
A-7	Ability to evaluate, analyze, and synthesize large quantities of data (which may be fragmented and contradictory) into quality, fused targeting/information products.	
A-23	Ability to understand and use the databases and tools to run queries to solve the business problem.	
A-15	Ability to identify information gaps.	
A-22	Ability to understand a business problem.	
A-24	Ability to identify patterns.	
A-17	Ability to recognize and mitigate deception in reporting and analysis.	
A-20	Ability to utilize multiple information sources across all information disciplines.	
A-2	Ability to identify basic common coding flaws at a high level.	Proficient in effectively using data analytics programming and visualization tools. Use data visualization tools to analyze data and produce reports. Complete the steps to design and implement a dashboard.
A-3	Ability to use data visualization tools (e.g., Flare, HighCharts, AmCharts, D3.js, Processing, Google Visualization API, Tableau, Raphael.js).	
A-5	Ability to communicate complex information, concepts, or ideas in a confident and well-organized manner through verbal, written, and/or visual means.	Competent in effective collaboration, communication and listening skills to define and solve business problem to a diverse audience. Develop algorithms using modular design principles to meet stated specifications. Identify, evaluate and suggest solutions to problems encountered in a team communication context.
A-9	Ability to effectively collaborate via virtual teams.	
A-13	Ability to adapt to a dynamic environment.	
A-14	Ability to function in a collaborative environment, seeking continuous consultation with other analysts and experts—both internal and external to the organization—to leverage analytical and technical expertise.	
A-21	Ability to effectively communicate ideas to team members with varying levels of technical expertise.	
A-18	Ability to think critically.	
A-16	Ability to recognize and mitigate cognitive biases which may affect analysis.	Maintain high standards of professional competence, conduct, and ethical practice.
A-11	Ability to exercise strong ethical judgment when policies are not well-defined.	

A-8	Ability to clearly articulate information requirements into well-formulated research questions and data tracking variables for inquiry tracking purposes.	Competent to identify and evaluate data relevance, reliability and validity from multiple sources to meet customer's needs. Research and utilize validated data to logically construct a report based on customer's needs.
A-12	Ability to focus research efforts to meet the customer’s decision-making needs.	
A-6	Ability to develop or recommend analytic approaches or solutions to problems and situations for which information is incomplete or for which no precedent exists.	
A-10	Ability to evaluate information for reliability, validity, and relevance.	
A-19	Ability to understand objectives and effects.	