GREAT with Data
Credentials Process for Data Science Area
DRAFT for US Department of Energy Consideration
August 31, 2020

The "GREAT with Data" initiative recommends developing two credentials for electric utility professionals in data science. The "GREAT with Data" also recommends that prospective employers of these professionals certify the credentials as meeting their objectives for employee proficiency.

1. Electric Utility Data Science "Leader": This credential is for leaders, managers and decision makers who need to understand the potential value and risks in electric utility data analysis, artificial intelligence and machine learning projects for allocating resources and overseeing results from work in this area. This credential does not require mathematics beyond algebra or programming expertise (although some supplemental material for those leaders who wish to try some programming will be included)
2. Electric Utility Data Science "Professional": This credential is for technical personnel, such as electric utility engineers, IT personnel, or data scientists entering the electric utility industry, who need to become data science professionals for an electric utility.

As nothing comparable currently exists for electric utilities, both are gaps that the GREAT with Data initiative will fill.

The terms "Leader" and "Professional" are draft in this context, and different terms may ultimately be used to distinguish the credential for those who will manage data science and those practicing data science. The final terminology will be developed as part of the training roadmap.

The "Professional" credential will be further sub-divided into either topical area (general practice, image processing, natural language processing, etc.) or by domain area (power generation, distribution), or potentially both. This sub-division will be developed as part of the training roadmap.

Each course should MINIMIZE passive time and MAXIMIZE active time. For "Leaders", this means lots of practical examples of immediate value. For "Professionals", this means a quick theory / objective lecture, followed by the practical problem and data to solve, including data importing, curation, cleaning, preprocessing, and model building.

The cross-cutting nature of data science should be emphasized throughout. For example, linear regression techniques for image processing are very applicable to Natural Language Processing and vice versa.

Electric Utility Data Science Leader

The table below has the proposed objectives for the "Electric Utility Data Science Leader" credential, as well an assessment on whether training material currently exists for the objective. This credential does not require mathematics beyond algebra or programming expertise (although some supplemental material for those leaders who wish to try some programming will be included). As noted above, the term "Leader" is provisional, and may be changed during development of the training roadmap.

Credential Objective: Apply knowledge of data analysis, artificial intelligence and machine learning to oversee project, conduct project analyses, and allocate resources in support of electric utility applications.

Technical Objectives

For each technical objective below, the following assessment is used:

1. No gap: there is existing courses and materials that can be used without modification
2. Small gap: there is existing courses and materials that can be used with minimal modification, typically to tie the existing material into the GREAT with Data credential
3. Gap: there is minimal existing courses and materials that can be used, and the GREAT with Data initiative will act to fill this gap.

Terminal Objectives	Gap?	Existing Material & Notes
1. Understand in data science such as machine learning, supervised, unsupervised learning, etc.	No or Small gap	Multiple on-line courses & articles; does a concise .pptx exist? "Intro to Data Analytics"
2. Understand data science ecosystem, such as where underlying libraries and programming tools are and require support	No gap	Is something specific for utilities needed? Is the existing material accessible to leaders? Note that this is a fast-changing area.
3. Apply steps to start an electric utility data science program, including required skills of personnel, numbers, equipment, software, etc.	Gap	Something specific for utilities needed?
4. Understand statements of accuracy and precision of algorithms	No gap	Might be part of "key concepts" objective above
5. Apply common sources of bias and errors in data science and to ask probing	No gap	Might be part of "key concepts" objective above

questions to determine how others have (or have not) dealt with bias		
6. Understand project management considerations for electric utility data projects including resource needs, cost-benefit assessments, and risk analysis.	Small gap	Not sure this exists anywhere yet?
7. Understand key concepts in electric utility data security and privacy protections	Small gap	Utilities have unique cyber requirements such as NERC-CIP, US NRC. However, also much existing training at utilities on these areas.
8. Correlate typical results in electric utility power generation application space, including: • Data Validation for thermal power efficiency • Image processing for inspection • Natural Language Processing to improve Business Process Automation	Gap	Derive from current ML101 course
9. Know common applications and typical results in electric utility transmission application space, including: • Transmission line losses	Gap	
10. Know common applications and typical results in electric utility distribution application space, including: • Resilience measurement • Reliability performance	Gap	
11. Know common applications and typical results in electric utility transmission application space, including: • Optimal power distribution	Gap	

12. Know common applications and typical results in general electric utility business application space, including: • Customer metrics • Supply chain.	Gap	

Electric Utility Data Science Professional

The table below has the proposed objectives for the "Electric Utility Data Science Professional" credential, as well an assessment on whether training material currently exists for the objective. This credential is for technical personnel, such as utility engineers or IT personnel, who need to become data science professionals for a utility. As noted above, the term "Professional" is provisional, and may be changed during development of the training roadmap.

As noted above, the "Professional" credential will be further sub-divided by either topic (general practice, image processing, natural language processing, etc.) or by domain (power generation, distribution), or potentially both. This sub-division will be developed in the training roadmap.

This credential and courses require mathematics through statistics & linear algebra, an expertise in an analytical programming language such as Python, R, etc., and familiarity with a visualization tool such as Power BI or Tableau. If an individual lacks the prerequisites, basic online resources will be recommended in:

1. Statistics
2. Linear algebra
3. At least one analytical programming language such as Python, R, C++ etc. for:
    a. Those with no or minimal programming experience
    b. Those who are adept in R and/or C++ but need to upgrade to Python for Deep Learning applications
4. Data visualization such as Power BI or Tableau

None of the above items are considered gaps, as there are multiple resources on-line and through common educational forums.

Credential Objective: Apply advanced knowledge of data analysis, artificial intelligence and machine learning to support electric utility applications at a scientist/engineering level.

Technical Objectives

For each objective, the following assessment is used:

1. No gap: there is existing courses and materials that can be used without modification
2. Small gap: there is existing courses and materials that can be used with minimal modification, typically to tie the existing material into the GREAT with Data credential
3. Gap: there is minimal existing courses and materials that can be used, and the GREAT with Data initiative will act to fill this gap.

Terminal Objectives	Gap?	Existing Material & Notes
1. Apply key concepts in data science such as machine learning, supervised, unsupervised learning, etc.	No or Small gap	Multiple on-line courses & articles; does a concise .pptx exist?

		"Intro to Data Analytics" **Could be skipped for an experienced data scientist.**
2. Understand & apply key concepts in the electric utility industry.	No or small gap	Multiple existing utility & EPRI materials; this is intended for data scientists entering the electric utility industry. Would not want to replicate utility on-boarding. **Could be skipped for an experienced data scientist.**
3. Operate data science ecosystem, such underlying libraries and programming tools are and require support	No gap	**Could be skipped for an experienced data scientist.**
4. Series of steps to start a utility data science program, including required skills of personnel, numbers, equipment, software, etc.	Small gap	Something specific for utilities needed?
5. Apply statements of accuracy and precision of algorithms	No gap	Might be part of "key concepts" objective above **Could be skipped for an experienced data scientist.**
6. Asses sources of bias and errors in data science and how to ask apply industry best practices to deal with bias	No gap	Might be part of "key concepts" objective above **Could be skipped for an experienced data scientist.**
7. Understand project management considerations for electric utility data projects	Small gap	Not sure this exists anywhere yet?
8. Apply key concepts in electric data security and privacy protections	Gap	Utilities have unique cyber requirements such as NERC-CIP, US NRC.
9. Collaborate with end-users to develop problem	Small gap?	Likely items on-line that can be utilized.

statement, data requirements and develop data science model for new applications not covered above		**Could be skipped for an experienced data scientist.**
10. Develop, apply and program models for common applications in electric utility power generation application space, including: 11. Data Validation for thermal power efficiency • Image processing for inspections • Natural Language Processing to improve Business Process Automation	Gap	Current ML101 course Each application may be a separate short course. Each would be like the "Leaders" course above, but with practical programming example and sample data / results.
12. Develop, apply and program models for common applications in electric utility transmission application space, including: • Transmission line losses.	Gap	Each application may be a separate short course. Each would be like the "Leaders" course above, but with practical programming example and sample data / results.
13. Develop, apply and program models for common applications in electric utility distribution application space, including: • Resilience measurement • Reliability performance	Gap	Each application may be a separate short course. Each would be like the "Leaders" course above, but with practical programming example and sample data / results.
14. Develop, apply and program models for common applications in electric utility transmission application space, including: • Optimal power distribution	Gap	Each application may be a separate short course. Each would be like the "Leaders" course above, but with practical programming example and sample data / results.
15. Develop, apply and program models for common applications in general electric utility business application space, including: • Customer metrics • Supply chain.	Gap	Each application may be a separate short course. Each would be like the "Leaders" course above, but with practical programming example and sample data / results.