Wichita State University

Master of Science in Data Science in Engineering

<u>Program Approval</u>

I. GENERAL INFORMATION

A. Institution Wichita State University

B. Program Identification

Degree Level:	Master's
Program Title:	Data Science
Degree to be Offered:	MS in Data Science
Responsible Department/Unit:	Electrical Engineering and Computer Science / College of Engineering
CIP Code:	30.7099
Modality:	Face-to-Face
Proposed Implementation Date:	Spring 2021

Total Number of Semester Credit Hours for the Degree: 30

II. CLINICAL SITES None

III. JUSTIFICATION

Data has been referred to as the "oil of the digital economy" due to its immense potential to optimal decision making ("The World's Most," 2017). **Data Science** (D.S.) deals with the generation of data, processing of it, and application and development of solutions using, but not limited to, machine learning, deep learning, and artificial intelligence. Big data is one of the most rapidly emerging topics in the world, creating a high demand for employees with expertise in various aspects of D.S. While this demand is global, due to the engineering and manufacturing nature of Wichita, it is especially important here.

Wichita State University (WSU) is strongly dedicated to supporting this data revolution by making teaching and research of data science and analytics an institutional priority. In support, WSU has recently invested in a High-Performance Computing (HPC) infrastructure and personnel. In this light, the MS in Data Science is one of the three distinct albeit aligned programs being proposed (the other two are in Business –MS in Business Analytics]; Liberal Arts and Sciences – MS in Mathematical Foundations of Data Analysis) to further this priority. All three share foundational coursework in business analytics, as well as other electives offered among the three colleges. The College of Engineering has elevated **computing and informatics** to a major priority and has made commitments for realigning and creating new programs in this field. In addition, WSU has chosen **Digital Transformation** as one of the pillars in *Convergent Sciences Initiative* and will be investing resources to help industry engage with academia in this space – data science is central to these efforts. Finally, the EECS Department has been allocated new resources to hire Dr. Dukka KC (director of the proposed program) as an Associate Professor to increase the D.S. capabilities at WSU and to create leadership potential in this space.

This proposed program offers students numerous opportunities to learn how to build a data pipeline and

transform raw data in ways that provide end-users a competitive advantage. Starting with a broad survey of data science and analytics, the bulk of the program focuses on the algorithmic and computing aspects of D.S. The integrated curriculum includes a capstone project focused on hands-on/experiential learning. This proposed program clearly advances WSU's mission to be an **essential educational, cultural, and economic driver for Kansas and the greater public good** by graduating students who are highly skilled in the sought-after field of data science.

Among public universities in Kansas, Kansas State University has an M.S. in Data Analytics program which is housed in College of Business and the University of Kansas has a M.S. in Applied Statistics and Analytics housed within the University of Kansas Medical Center. Although both of these programs help to fill some of the gap in the number of professionals with these skills, the available/unfilled D.S. related jobs are ever-increasing. In addition, the proposed program is unique as the focus is to develop data scientist and engineers who are well versed in algorithmic and computational thinking to develop data science related tools and infrastructure. Moreover, the majority of WSU students are from (or within 30 miles of) Wichita, thus locating a graduate program in Data Science at WSU provides the educational opportunity for students in the region.

IV. PROGRAM DEMAND

A. Survey of Student Interest

Number of surveys administered:	250
Number of completed surveys returned:	100
Percentage of students interested in program:	81%

The survey was sent to 250 undergraduate and graduate students in the Department of Electrical Engineering and Computer Science (EECS) at WSU to inquire about the need for a master's degree in Data Science program. 57% percent of the respondents are undergraduate students while 43% are graduate students. Among these students, almost half (51%) were already taking some data science-related courses. Among the respondents, 91% saw significant value in a data science program at WSU. Similarly, 82% of the respondents said that they would consider enrolling in an M.S. D.S. program if it were offered. Finally, 87% of the respondents said that they would likely recommend this program to their friends.

B. Market Analysis

The *Harvard Business Review* calls data science the "**sexiest job of the 21st Century**" (Davenport & Patil, 2012). In addition, there is a significant demand for professionals with data science skills. Various reports and reviews have consistently pointed out the large gap in the number of professionals with these skills and available jobs in this area. Moreover, there has been a steady increase in the employment of data scientists, but demand is expected to grow even faster. According to Glassdoor, data scientist was the top job in America for the second year in a row in 2017 (Junco, 2017). In addition, Glassdoor cites that top among the benefits of a career in data science is a median base salary of $110,000 and a knowledge base that is applicable to practically any field. This report also states that overall job satisfaction that comes with being a data scientist ranks 4.4 out of 5 dominating over several other highly-sought-after careers for the title of "best job." The field of data science is experiencing rapid growth as new technology is developed and more data becomes available.

Data science growth is only expected to continue to develop and expand in the future. In fact, the Bureau of Labor Statistics (BLS), which reports employment data throughout the United States, has projected a 31% increase for statisticians and data scientists by the year 2026. BLS is not the only entity highlighting this as an essential current need. The same report from Glassdoor mentions that seven of the top ten spots are related to information technology (IT), and four are related to data management, including data engineer,

analytics manager, database administrator, and mobile developer. In fact, recent data from job sites show that there has been a 29% increase in demand for data scientists year after year and a 344% increase since 2013 (Flowers, 2019).

A recent report from the American Statistical Association (ASA) highlights the continued growth and demand for graduates with data science and analytical skills (2019). LinkedIn recently highlighted the fact that data science dominates the ranking of emerging jobs searched/available (Dignan, 2019). Of potential concern, searches by job seekers skilled in data science grew at a slower pace (14%), suggesting an increasing gap between supply and demand, which may be partially due to D.S. skills being typically acquired via an M.S. degree (Kolakowski, 2020). In response, D.S. M.S. programs are being developed rapidly across the country, and will soon become as critical as e.g., biology or psychology programs. Finally, the need is such, that a few of our own EECS graduates have already been hired as data analysts/scientists without having formal extensive education in data science. In sum, we are very excited about the job prospects for our future D.S. graduates.

V. PROJECTED NEW ENROLLMENT FOR INITIAL THREE YEARS OF PROGRAM

Year	**Headcount Per Year**		**Semester Credit Hours Per Year**	
	Full-Time	**Part-Time**	**Full-Time**	**Part-Time**
Implementation	15	0	270	0
Year 2	30	5	720	45
Year 3	30	5	900	90

VI. EMPLOYMENT

A Bureau of Labor Statistics (BLS) report, “Occupational Outlook Handbook, Mathematicians and Statisticians,” as well as other private sector reports, “Data Scientist: A Hot Job that Pays Well” (Flowers, 2019), and “New Report Highlights Growing Demand for Data Science, Analytics Talent, Steps for Higher Ed and Business Recommended” (ASA, 2019), clearly demonstrate the sharply increasing gap between the need for students with data science-related degrees and the supply of these students. In addition, the employment market for data scientists is robust, with a growing need for qualified data scientists/engineers. Through the aforementioned surveys and reports, it has also been well established that this need spans a variety of industries including technology, finance, telecommunications, manufacturing, service, retail, banking, cybersecurity, and others (Smith Hanley Associates LLC, 2018). Critically, the BLS also shows that Wichita, Kansas, is one of the metropolitan areas with a high demand for such jobs.

According to a recent jobs report by Glassdoor, based on the number of job openings, salary, and overall job satisfaction, data scientist is ranked number one, with more than 4,000 job openings; data engineer is ranked number three, with more than 2,500 job openings; and analytics manager is ranked number five, with almost 2,000 job openings (Junco, 2017). In addition, the number of data scientists has more than doubled over those five years, and the number of data engineers sextupled. Training in data science is relevant to many job titles, including statistician, computer systems analyst, software developer, database administrator, computer network analyst, data scientist, data analyst, data engineer, and data manager.

The proposed program with its state-of-the art curriculum and the inclusion of a real-world capstone will enable students to develop an array of competitive skills that will enable them to pursue a wide range of data science career paths. Some of the potential employment opportunities for graduates with a master’s in D.S. include data scientist, data engineer, business intelligence specialist, data analyst, and others.

VII. ADMISSION AND CURRICULUM

A. Admission Criteria

Students will be admitted in full graduate standing in the M.S. in Data Science program if they have a bachelor's degree in computer science or any related engineering discipline and a GPA of at least 3.00, and also meet the Graduate School's other requirements. Students who have a bachelor's degree in other quantitative disciplines (Mathematics, Physics, or other disciplines) with demonstrated quantitative skills (calculus, linear algebra, etc.) and proficiency in computer programming may be admitted on a conditional basis.

Application materials will be reviewed by the Graduate School and the Data Science graduate coordinator, after which the student will be notified of their decision. Students entering the M.S. in Data Science program are expected to have already completed courses in programming, linear algebra, statistics, and data structures. If prior coursework deficiencies exist, then the student may be admitted on a conditional basis. It is recommended that deficiencies are completed prior to beginning graduate studies.

B. Curriculum

The proposed program emphasizes development of the next generation of data scientists and engineers. Students graduating from the program will master the skills to build the infrastructure for delivering insights from raw data sources, as well as implement data science pipelines and workflows for acquiring, cleaning, transforming, analyzing, and visualizing data to provide descriptive, predictive, and prescriptive analytics. The program includes a curriculum to develop sought-after skills in various aspects of data science and engineering to prepare a skilled workforce in the area of data science.

The overall objectives of the proposed M.S. in Data Science program are to ensure that a graduating student possesses the following:

- Technical knowledge on data science principles, computational tools and algorithms, data science life cycle, data-driven problem-solving process, and management of data and information to solve data-intensive problems and to describe and transform data to knowledge/information.
- Effective communication and technical knowledge in cleaning, processing, analyzing data and effective visualization so that they are able to communicate solutions to stake holders and broader audience.
- Knowledge of modern machine learning techniques and data science tools and software skills to build predictive and analytical workflows.

To achieve these objectives, the curriculum will consist of 30 credit hours, including core courses that all students must complete, computer science (CS) elective courses, and other elective courses (cf. **Table 1**). Students must complete 15 credit hours of core courses that will provide sufficient background in data science, including Data Science, Mathematical Foundation of Data Science, Machine Learning, and Business Analytics. The curriculum will also require each student to complete a Capstone Project in Data Science course. They must also complete nine credit hours of elective coursework in CS, and six hours of elective coursework from other related disciplines. Depending upon the student's background, all 15 elective credits may be obtained from CS electives. Particularly noteworthy is that nine credits (one course per) are shared between the three programs being proposed together, **giving students a unique "bird's eye" view of the full data science and analytics space – from theory to practice to business implementation**.

The curriculum requires 30 hours for graduation, and students must earn a 3.0 overall GPA. One of the salient features of the MS curriculum is that each student must take the Capstone Project in Data Science course. Students in this course will engage in all data science life-cycle process topics including data collection, preprocessing, transformation, exploratory data analysis, visualization, predictive modeling, descriptive modeling, clustering, regression and classification, and data science project life cycle. The project topic will come from an academic research group, industry, government, other stakeholders, or other sources that mimic a real-world data science problem. Please refer to the syllabus of the capstone course for details about this. Almost all the courses are existing courses besides the capstone course.

Table 1: MS in Data Science Program Course Schedule

Year 1: Fall **SCH = Semester Credit Hours**

Course #	Course Name	SCH….
CS 697AK	Introduction to Data Science	3
BSAN 775**	Perspectives on Business Analytics	3
MATH 746**	Introduction to Data Analytics	3

Year 1: Spring

Course #	Course Name	SCH….
CS 697AB	Machine Learning	3
DSE	Data Science Elective Course (see Table 2)	3
DEC/DSE	Discipline Elective Course or Data Science Elective Course (see Table 2)	3

Year 2: Fall

Course #	Course Name	SCH….
DSE	Data Science Elective Course (See Table 2)	3
DSE	Data Science Elective Course (See Table 2)	3
DEC/DSE	Discipline Elective Course or Data Science Elective Course (See Table 2)	3

Year 2: Spring

Course #	Course Name	SCH….
CS 896**	Capstone Project in Data Science	3

** represents new course

Total Number of Semester Credit Hours ……. [30]

Table 2: MS in Data Science Program Courses

Course No.	Course Name	Credits
Five Required Courses—15 Credits		
CS 697AK	Introduction to Data Science	3
BSAN 775**	Perspectives on Business Analytics	3
MATH 746**	Introduction to Data Analytics	3
CS 697AB	Machine Learning	3
CS 896**	Capstone Project in Data Science	3
Three Data Science Elective Courses (DSE) —9 Credits		

Course No.	Course Name	Credits
CS 665	Introduction to Database Systems	3
CS 771	Artificial Intelligence	3
CS 797I	Introduction to Bioinformatics	3
CS 898AS	Deep Learning: Theory, Algorithms and Applications	3
CS 898AJ	Big Data Analytics	3
CS 898BE	Advanced Topics in Machine Learning	3
CS 898CA	Introduction to Intelligent Robotics	3
CS 898BA	Image Analysis and Computer Vision	3
CS 898AW	Artificial Intelligence for Robotics	3
CS 898D	Data Mining	3
CS 898BD	Deep Learning	3
Two Other Discipline Elective Courses (DCE) —6 Credits		
MIS 750	Business Intelligence and Analytics	3
STAT 763	Applied Regression Analysis	3
STAT 764	Analysis of Variance	3
STAT 776	Applied Statistical Methods	3
IME 780AP	Neural Networks and Machine Learning	3
IME 869	Bayesian Statistics and Uncertainty Quantification	3
SMGT 800	Analytics and Decision Making in Sport	3
IME 780AN	Big Data Analytics in Engineering (if CS 898AJ not taken)	3
IME 734	Introduction to Data Mining and Analytics (if CS 898D not taken)	
MIS 884	Database Management and Planning	3
BSAN 875	Advanced Business Analytics	3

** represents new course

VIII. CORE FACULTY

Faculty Name	Rank	Highest Degree	Tenure Track (Y/N)	Academic Area of Specialization	FTE Devoted to Proposed Program
Dukka KC*	Assoc. Professor	Ph.D	Y	Data Science/Bioinformatics	30%
Rajiv Bagai	Professor	Ph.D.	Y	Web Anonymity, Deductive Databases	10%
Sourabh Bose	Asst. Professor	Ph.D.	N	Machine Learning	10%
Hongsheng He	Asst. Professor	Ph.D.	Y	Intelligent Robotics	10%
Vinod Namboodiri	Professor	Ph.D.	Y	Mobile Computing	5%
Ajita Rattani	Asst. Professor	Ph.D.	Y	Biometrics, Computer Vision	20%
Sergio Salinas	Asst. Professor	Ph.D.	Y	Privacy and Security	5%

Kaushik Sinha	Assoc. Professor	Ph.D.	Y	Machine Learning/Data Mining	15%

*Director of Program

FTE: 1.0 FTE = Full-Time Equivalency; In FTE we also consider that all tenure track faculty have responsibilities for service and research in addition to teaching.

Number of Graduate Assistants Assigned to this program: 2 - 7

IX. Expenditure and Funding Sources

A. EXPENDITURES	First FY	Second FY	Third FY
Personnel—Reassigned or Existing Positions			
Faculty	$103,931	$107,049	$213,261
Administrators *(other than instruction time)*	$14,931	$15,379	$15,840
Graduate Assistants	$19,200	$19,776	$40,145
Support Staff for Administration (*e.g., secretarial*)	$7,665	$7,895	$8,132
Fringe Benefits *(total for all groups)*	$44,154	$45,479	$83,305
Other Personnel Costs	--	--	--
Total Personnel Costs—Reassigned or Existing	**$189,881**	**$195,578**	**$360,683**
Personnel—New Positions			
Faculty	--	$95,000	$95,000
Administrators *(other than instruction time)*	--	--	--
Graduate Assistants	--	$19,776	$33,990
Support Staff for Administration (*e.g., sec*	--	--	--
Fringe Benefits *(total for all groups)*	--	$31,977	$33,399
Other Personnel Costs	--	--	--
Total Personnel Costs—New Positions	**$0**	**$146,753**	**$162,389**
Start-Up Costs—One-Time Expenses			
Library/learning resources	--	--	--
Equipment/Technology	--	$60,000	$60,000
Physical Facilities: Construction or Renovation	--	--	--
Other	--	--	--
Total Start-Up Costs—One-Time Expenses	**$0**	**$60,000**	**$60,000**
Operating Costs—Recurring Expenses			
Supplies/Expenses	$2,000	$3,000	$4,000
Library/Learning Resources	--	--	--
Equipment/Technology	--	$12,000	$24,000
Travel	--	--	--
Other	--	--	--
Total Operating Costs—Recurring Expenses	**$2,000**	**$15,000**	**$28,000**
GRAND TOTAL COSTS	**$191,881**	**$417,331**	**$611,072**

B. FUNDING SOURCES (projected)	**Current**	**First FY (New)**	**Second FY (New)**	**Third FY (New)**
Tuition/State Funds		$143,688.60	$397,028.70	$506,950.20
College Course Fees		$13,500	$38,250	$49,500
Student Support Fees		$20,375.40	$62,258	$83,765.20
GRAND TOTAL FUNDING		**$177,564.00**	**$497,536.70**	**$640,215.40**

C. Projected Surplus/Deficit (+/–) (Grand Total Funding *minus* Grand Total Costs)		($14,317.00)	$80,205.70	$29,143.40

X. EXPLANATIONS OF EXPENDITURES AND FUNDING SOURCES

A. Expenditures

Personnel—Reassigned or Existing Positions

- **For the first year**:
  - The table from Core Faculty was taken and salaries are multiplied by the FTE. Note that this merely represents a slight reorganization as this cost is latent.
  - Administrator is calculated as 8% of the Chair's salary (again, a latent cost).
  - Two GTAs are calculated for the 270 credit hours.
  - Support staff cost is calculated as 25% of the current graduate secretary's salary (latent cost).
  - Fringe is calculated based on the current WSU fringe rates.
- **For the second year**:
  - All costs are increased by an estimated 3% of raise.
- **For the third year:**
  - All costs are increased by an estimated 3% of raise.
  - Expenses from the second year *new* are carried to the third year *existing* and increased by 3%.

Personnel—New Positions

- **For the first year**:
  - The estimated 15 student enrollment can be managed by existing resources as most of the costs are latent.
- **For the second year**:
  - We estimate the enrollment to increase by 30 students, meaning we should consider adding an additional faculty member. The cost in this case is 100% of the new faculty's salary as this program will be the reason to hire that faculty.
  - The increase in credit hours also requires 2 more GTAs (with a very lean 200CH/GTA estimate)
- **For the third year:**
  - We estimate the enrollment to increase by an additional 30 students, meaning a second new faculty member should be considered. Again, the cost is 100% of the new faculty's salary as this program will be the reason to hire this person.
  - The increase in credit hours will require an additional 3 GTAs (with a very lean 200CH/GTA estimate – averaging for this number to be an integer)
  - All costs are increased by an estimated 3% of raise

Start-Up Costs—One-Time Expenses

- **For the first year:**
  - Current resources are sufficient.
- **For the second year:**
  - 30 computers for a lab are included (at $2k each) to keep up with student population growth.
- **For the third year:**
  - An additional 30 computers for a lab are included (at $2k each) to keep up with continued student population growth.

Operating Costs—Recurring Expenses

- **For the first year:**
  - Supplies (copying, office supplies) are estimated at $2k.
- **For the second year:**
  - Supplies are estimated at $3k.
  - 30 computers amortized over 5 years result in $12k latent cost.
- **For the third year:**
  - Supplies are estimated at $4k.
  - 60 computers amortized over 5 years result in $24k latent cost.

B. Revenue: Funding Sources

Revenue is calculated based on the enrollment table from Section V:

- $307.98 graduate tuition is calculated for half the full-time student credit hours.
- $756.38 international graduate tuition is calculated for the other half of full-time student credit hours since are expecting a large portion of the D.S. students to be international.
- All part-time student credit hours are calculated using the $307.98 tuition.
- Student fee:
  - 7.00 or more credit hours $679.18/fall or spring semester
  - 4.00-6.75 credit hours $452.78/fall or spring semester
- $50 per credit is added to all credit hours for the program fee applied to all College of Engineering programs.

C. Projected Surplus/Deficit

- While the first year shows a deficit, this is a latent cost (which is far outweighed by the opportunity cost of not developing the program). Most costs are covered through current resources with a small reorganization.
- With the addition of two faculty, we expect the program to grow, with healthy surpluses over time, which will allow for additional paid GTA's.

XI. REFERENCES

American Statistical Association. (2019). New report highlights growing demand for data science, analytics talent, steps for higher ed and business recommended. American Statistical Association. https://www.amstat.org/ASA/News/New-Report-Highlights-Growing-Demand-for-Data-Science-Analytics-Talent.aspx

Davenport, T.H. and Patil, D.J. (2012, October). Data scientist: The sexiest job of the 21st century. *Harvard Business Review*. https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century

Dignan, L. (2019, Dec. 10). Data science dominates LinkedIn's emerging jobs ranking. ZDNet. https://www.zdnet.com/article/data-science-dominates-linkedins-emerging-jobs-ranking/

Flowers, A. (2019, Jan. 17). Data scientist: A hot job that pays well. Indeed Hiring Lab. https://www.hiringlab.org/2019/01/17/data-scientist-job-outlook/

Junco, P.R. (2017, Sept.21). Data scientist personas: What skills do they have and how much do they make? *Glassdoor Economic Research*. https://www.glassdoor.com/research/data-scientist-personas/

(Kolakowski, N.) (2020, Feb. 13). Fastest-growing tech occupations include data scientists, engineers. Dice Insider. https://insights.dice.com/2020/02/13/fastest-growing-tech-occupations-data-scientist-engineer/

Smith Hanley Associates LLC. (2018, May 24). Top data scientist utilization by industry, sector, function. Smith Hanley Associates LLC. https://www.smithhanley.com/2018/05/24/data-scientist-industry-sector-function/

The world's most valuable resource is no longer oil, but data. *The Economist*. May 6, 2017 https://www.economist.com/leaders/2017/05/06/the-worlds-most-valuable-resource-is-no-longer-oil-but-data

U.S. Department of Labor. (n.d.). Occupational Outlook Handbook, Mathematicians and Statisticians. Bureau of Labor Statistics. https://www.bls.gov/ooh/math/mathematicians-and-statisticians.htm