A Proposal for a Professional Master's Degree Program in Computational Data Analytics at New Mexico State University

Administered by the College of Arts and Sciences, the College of Business, the College of Engineering, and the Graduate School

Las Cruces, New Mexico

Contact person who can answer specific questions about the program:

Dr. Enrico Pontelli
Dean
College of Arts & Sciences
New Mexico State Univ.
epontell@nmsu.edu
(575) 646-3500

Dr. Huiping Cao
Associate Professor
Dept. Computer Science
New Mexico State Univ.
hcao@cs.nmsu.edu
(575) 646-4600

Dr. Loui Reyes
Dean
The Graduate School
New Mexico State Univ.
louireye@ad.nmsu.edu
(575) 646-5745

1 Executive Summary

The proposed Professional Master Computational Data Analytics (PMCDA) program is designed for students interested in developing expertise in data analytics, with specialization in computational analytics. Data analytics is an inherently interdisciplinary discipline, dealing with methods and systems to synthesize knowledge or insights from large quantities of data collected from heterogeneous sources and diverse spatial and time scales. Data analytics employs theories, methodologies, and tools drawn from many fields, within the broad areas of mathematics, statistics, and computer and information sciences, and applies them to a diversity of data-rich domains, such as life sciences, medicine, physical sciences, social sciences, engineering, business, and education.

Data analytics is a very broad and multifaceted field. The *Computational Analytics* emphasis will enable students to address analytics problems that are associated to *large* quantities of data, to discover new knowledge and support decision making. To accomplish this, the PMCDA program will provide students with a strong foundation in data management and analysis, the computational and statistical thinking, and understanding of computer systems. The program will provide its graduates with the necessary skills to succeed in data science-related jobs, which require the integration of data analysis skills, ability to deal with large quantities of data, and a solid foundation in computing. The professional focus of the degree will prepare students for success in the workplace, with an emphasis on enriching the preparation of students who are already in the workplace and are seeking technical skills to advance their careers in the data analytics domain.

After completing the PMCDA program, students will have gained the skills and ability to:

- Analyze real-life data from diverse sources and domains
- Effectively apply analytics tools to large data sets
- Apply mathematical and statistical models to data analysis problems
- Apply computational thinking to develop effective data analytics solutions
- Apply programming and debugging skills to problem solving
- Understand and use computer technology and software in solving real-life data analysis problems
- Understand and address unfamiliar problems related to data analytics
- Develop effective instrument to communicate solutions to diverse audiences

2 Purpose and Mission of the Program

2.1 Overview of the Proposed Program

The terms *"big data"* and *"data analytics"* are frequently used to describe the challenges, techniques and methodologies used to deal with the vast amounts of data currently being generated and analyzed in many areas of science, education, business and government. In 2013, IBM estimated that over 2 million terabytes of data are created every day - equivalent to over 300 million high definition movies. The National Security Agency gathers as much information as is stored in the Library of Congress every 6 hours. Over 90% of the world's data was generated in just the past two years.

Making sense of this vast sea of data for the use and benefit of society is an imperative; indeed higher education institutions and companies are already strategizing and restructuring for this "big

data" tsunami. The ability to extract information from these data sets, and in turn convert the information into knowledge, has already led to ground-breaking advances in a wide variety of domains, ranging from genomics, to financial trading, to high-energy physics and education. Data analytics is supporting the development of new information-based industries – according to a recent report from International Data Corporation (IDC), this represents a market estimated at $130 billions and expected to grow to over $203 billions by 2020 [6].

Techniques to support big data are traditionally grounded in the field of *computing* – computational methods are essential to process the data sets whose size is in the order of terabytes and beyond. Until recently, methods of analysis of these data had largely assumed that "traditional" computing inferential tools are adequate to summarize and make informative inferences from the data. With the increased capacity to collect, store and summarize data, and with the growth in the level of sophistication of the questions that we seek to answer, it is becoming clear that data analytics requires skills that go beyond the mere application of tools, skills that can be developed through specialized training in data management and data analytic skills. Data analytics and data science have emerged crucial interdisciplinary disciplines for developing solutions for gathering, cleaning, archiving, analyzing, usingm and visualizing data for the purposes of making informed decisions.

A popular report by McKinsey Global Institute [8] estimates that, by 2018, the US could face a shortage of close to 200,000 people with deep analytic skills, and 1.5 million analysts will be needed with the expertise to translate the analysis of available data into actionable decisions. The "deep analytic skills" expected are primarily driven from tools, skills, methodologies, and techniques that come from the disciplines of Computer Science, Mathematics and Statistics.

A recently published report [2] by Burning Glass Technologies, Business-Higher Education Forum (BHEF), and IBM states that "*the demand for data scientists will soar 28% by 2020*". A report at Burtch Works, a leading executive recruitment agency that matches top Big Data & Data Science talent with top jobs [4] emphasized three key trends that shape 2017 data science hiring. Among these three key trends, one of them is "*More early career data scientists opting for Master's degrees, not PhD's.*"

The mission of the proposed degree program is to produce trained professional with deep analytic skills, by drawing on these core disciplines. The core courses of the Professional Master in Computational Data Analytics (PMCDA) program come primarily from statistics and computer science, but the program also recognizes that some of the skills necessary for big data and data analytics come from the needs and applications in specific domains; the program includes advanced courses and culminating experiences that enable students to apply their analytics skills to problems in a specific domain.

2.2 Program Description

We propose this new degree program in response to this exploding "big data" industry sector. New Mexico State University is well-positioned to deliver such a degree because of existing resources - including faculty members (e.g., the solid expertise of the Applied Statistics department, the focused hires conducted in Computer Science), computer equipment, curriculum and courses – within existing colleges and departments.

The program provides a progressive exposure to computational data analytics:

- The **foundational layer** provides training in the statistical and computational skills that underlie data analytics

- The **methodology layer** explores the tools, techniques and methodologies that are commonly employed in data management, data mining and machine learning
- The **advanced layer** provides a range of electives to explore advanced methodologies in data analytics and their use in diverse domains
- The **capstone experience** allows students to apply their skills and knowledge to solve a real-world problem

2.3 Academic Objectives

The primary academic purpose of this professional degree program is to develop professional workforce that is prepared to address the needs in the rapidly expanding computational data analysis field. The students will be educated to work and excel in a variety of work settings, including private corporations, national laboratories, government and educational settings. By educating these students we will be supplying qualified workforce to an industry which is currently looking for cost effective ways of delivering more effective decision making based on diverse and heterogeneous data sets.

The primary academic objectives for this proposed new degree program will be the development of high level skills in:

- Identifying and defining problems and decisions that can be answered by data
- Acquiring, analyzing and exploring data
  - Acquiring: getting, cleaning, archiving, integrating data
  - Analyzing: visually, mathematically, statistically
  - Exploring: seeking trends and patterns
- Managing and communicating data narratives (stories) that transform data into actionable information
- Exposure to real-world problems, through applied course in different disciplines and exploring partnerships with industry and national labs involving big data and data analysis.

2.4 Program Curriculum

2.4.1 Curriculum

The admission requirements for the degree program will require incoming students to have a minimum mathematical preparation at the level of *Linear Algebra* (MATH 280 or equivalent course, such as E E 200).

The curriculum for the degree program is composed of 34 graduate credits. The degree requirements are articulated in three phases:

1. **Foundational:** The background courses are aimed at developing the necessary computing and mathematical background to operate in the domain of data analytics. The following courses are required:
   a. **Programming Foundations: [6 credits]**
      - Programming R:
        - CS 458 or A ST 515 for students with low computing background
        - CS 459 for students with existing computing background
      - Programming in Python:
        - CS 453 for students with low computing background
        - CS 454 for students with existing computing background

b. Mathematical Foundations: [7 credits]

- Statistical Inference and Regression:[1]
    1. A ST 505 and
    2. A ST 507

2. **Methodologies: [9 credits]** The foundation courses provide the preparation in the basic methodologies and technologies needed for data analytics.
    a. Database Management Systems: ICT 458 or CS 502 or BCIS 575
    b. Data Mining: CS 508 or A ST 550
    c. Machine Learning: CS 519 or EE 565
3. **Advanced Topics and Applications: [9 credits]**
    a. Multivariate Analysis: A ST 555
    b. Advanced Mathematical Modeling: MATH 518, STAT 535
    c. Data Analysis: IE 590, IE 545, BCIS 561
    d. Modeling and Simulation: CS 512, IE 567, IE 515, IE 522
    e. Communication: COMM 550, ENGL 543
    f. Visualization: CS 506, ICT 460
    g. Predictive Analytics: BCIS 566
    h. Enterprise Systems: BCIS 585
    i. Advanced Data Management: CS 582
    j. Applications:
        - Astronomy: ASTR 630
        - Biology: BIOL 566
        - Electrical Engineering: EE 596
4. **Capstone Experience: [3 credits]**
    a. Internship, or
    b. Master's Project or Thesis: CS 598 or EE 598 or IE 599 or MATH 599 or A ST 598

2.4.2 Learning Outcomes

Upon successful completion of this program, graduates will be well positioned to find employment in the burgeoning data science and analytics industry. This will be an interdisciplinary degree drawing on faculty from various colleges (Arts and Sciences, Business, Engineering) and programs (e.g., applied statistics, mathematical sciences, computer science, industrial engineering, electrical and computer engineering, astronomy, criminal justice).

Learning outcomes for each of the individual courses will be provided by the instructors of the course, who will have the primary responsibility to see that these outcomes are achieved. The overall goal of the Computational Data Analytics programs is to equip quantitative professionals with the tools to gather, analyze and interpret data collected on scales up to terabytes (10^{12} bytes) or even petabytes (10^{15} bytes). Upon completion of the PMCDA program, students will be able to:

1. Define problems to be addressed through data analysis techniques

[1] Students with prior training in statistical methods through advanced regression may opt to replace with the sequence STAT 470 and STAT 480 or with the sequence A ST 565 and A ST 566.

2. Describe and apply the basic principles of statistical inference and commonly used statistical models, data mining, machine-learning tools, and database tools.
3. Select appropriate statistical and computational methods and apply analytical skills to effectively summarize, visualize, and make valid inferences from data.
4. Describe, analyze, apply, and use data bases, data mining and machine learning tools.
5. Fit and evaluate simple and multivariate linear regression models and generalized linear models.
6. Perform appropriate analyses on time-series data and multidimensional data.
7. Show proficiency in R, Python, and database programming.
8. Communicate quantitative results to individuals who may not have expertise in either statistics or computer science.
9. Apply essential database skills such as storing, organizing and manipulating large amounts of data contained in databases.
10. Perform data collection, cleaning (e.g., harmonize, rescale, parse, convert) and data imputation.
11. Articulate principles of statistical and data ethics.

Although students will undoubtedly learn other technologies (e.g., Excel, HTML, CSS, JavaScript) within this master's degree they will become expert in the following technologies since these are pervasive in the data science industry:

- An operating system of choice for data science (initially this would be Linux)
- A high level scripting language for quantitative data manipulation and visualization (initially Python)
- A high level programming tool for statistics and machine learning (initially R)

These Learning Outcomes will be assessed with a focus on coursework learning outcomes, final capstone project, data analysis outcomes, testing, conformance to best practices, student portfolios, class presentations and a variety of evaluations. Exit surveys will be carried out with all students upon finishing the program.

3 Program Justification

3.1 Need for the Program

The McKinsey Report suggests that by 2018 in the USA the data science industry will face a shortage of professional workforce of close to 200,000 people. A search of any of the large job websites (such as simplyhired.com) reveals hundreds of current openings in data science.
In terms of compensation, recent studies [7, 3] report:

- Data scientists who are managers make considerably more than those who are individual contributors, but for both, compensation increases significantly with scope of responsibility and years of experience.
- The median base salary is between $87K and $91K. The median salary raises to $106K if we focus on US jobs only.
- Salaries are changing at a rapid pace; a 2016 study indicates that half of the sample witnessed a 20% change, and the salary of 12% of the sample doubled.
- The highest median salaries belong to those who code 4–8 hours per week; the lowest to those who don’t code at all.

Another telling sign regarding the need for this type of program is the nationwide proliferation of graduate programs, primarily at the Master's level, in data analytics, data science, and programs with similar titles, aimed at preparing professionals to deal with large and diverse data sets. Some of these programs have acceptance rates below 15%. In response to the increasing demands on techniques and skills of big data analytics, almost every US state has supported Master's degrees in different universities [1]. However, according to Master's in Data Science [1], in New Mexico, there is no official Master's degree on big data and data analytics.
Enrollment in relevant courses at NMSU has been growing steadily over the years. The overall enrollment in the relevant AStat and CS courses has increased from 119 in Fall 2013 to 152 in Fall 2016.
A market analysis conducted in 2013 by the Department of Statistics of Oregon State University found that about 90% of students in certificate programs focused on Data Analytics were currently employed, in professional, scientific, technical services, educational services, insurance providers and health organization. Many of such students are pursuing such studies to acquire new skills to transition into a new functional area. Data Analytics has gained traction over traditional statistics and applied mathematics programs; at Colorado State University, for example, there were twice as many inquiries regarding "data analytics" as for "applied statistics." The popularity of the concepts can be seen even in simple popularity web searches (Figure 1).

According to a very recent ACM Computing Survey on data science [5], about 500 subjects of courses are related to data science or analytics. Among these courses, about "*72% are offered at Master's level, with only 7% at bachelor level, and 3.6% at the doctoral level*". It shows that a Master's degree on data analytics could be easily fit into the current educational system. When it comes to the media of offering such courses, online courses significantly complement the traditional education systems. For example, there are increasing number of courses offered in the Massive Open Online Course (MOOC) mode such as Coursera and Udemy.

Data science can be applied to almost every area that generates, collects, analyzes, and stores data. This program will prepare students working in the application areas that demand for professional skills of big data analytics. In particular, such application areas include biotechnology, energy, finance, gaming and hospitality, government, health care, insurance, internet, manufacturing, pharmaceuticals, retail, telecommunications, travel and transportation, and utilities. We can summarize the top skills required for data science positions by going through the requirements for such positions (at different levels) at different hiring websites. The top skills include machine learning, data mining, statistical analysis, Python, R, Support Vector Machine (SVM, one type of machine learning algorithms), k-Nearest Neighbor (kNN, one type of data mining/machine learning algorithms), MapReduce, and SQL. The levels of the positions generally depend on the years of work experience in the area and the levels of degrees. Most entry-level positions require 0-2 years of work experience in those areas and a Master's degree. While the middle-level or senior positions require more years of experience or higher degree (Master's or doctoral degrees). Given the requirement from employers at the hiring websites, our program will train students who would like to get entry-level or middle-level data science positions by covering the majority of the required professional skills.

The research results of different reports and surveys also echo the necessity of a canonical set of skills and techniques. Most online courses cover the classical subjects such as statistics, data

mining, machine learning, prediction, database management, and newer knowledge including analytical languages such as R and Python, and cloud infrastructure MapReduce and Hadoop [5]. As shown in Figure 2, the MGI 2016 report [9] lists the major techniques, which are machine learning centered and are combined with other analytical skills.

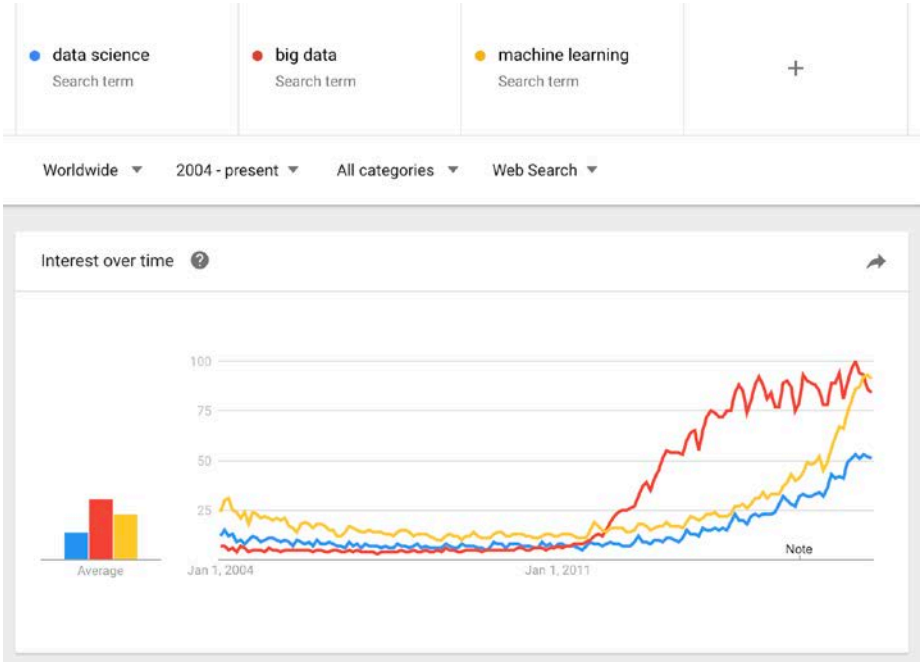


Figure 1: Popularity of Data Science

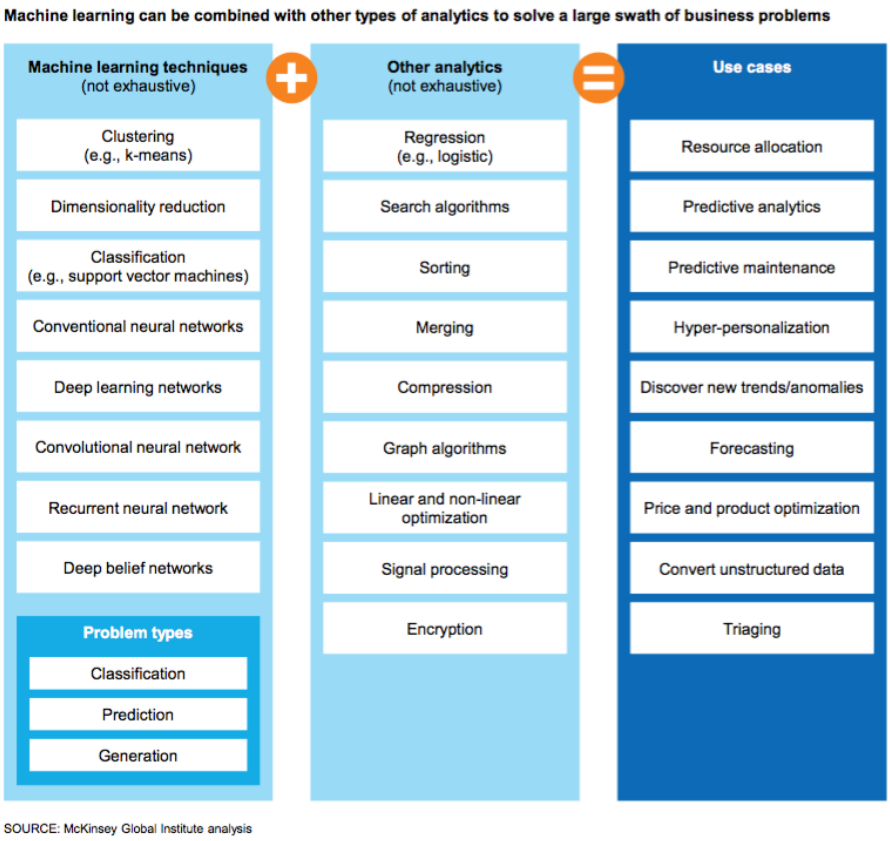


Figure 2: Analytics Skills

3.2 Relationship to NMSU Mission

The proposed program will have direct relevance to several of the Vision 2020 goals and values. The Vision 2020 document [10] lists the following core values:

1. Diversity and Inclusion: the proposed degree program will build on a wealth of expertise in the area of broadening participation and accessibility in STEM and computing. The proposed degree program will be accessible to a broad audience with diverse backgrounds, and it will contribute to the diversification of the workforce in the area of data analytics.
2. Accountability: Precise quantitative and qualitative metrics will be established to provide effective and ongoing monitoring of the program. Separate metrics will be developed to:
    a. Assess implementation and deployment of the degree program
    b. Assess learning outcomes of students, acquisition of critical data analytics skills, and students' access to rewarding careers relevant to computational data analytics.
3. Excellence: The program will meet the highest educational standards, reflecting the cutting-edge in principles, methodologies and technologies in the areas of computational data analytics. The program will contribute to promoting the educational and research profile of the institution.
4. Discovery: The degree program will build on a range of ongoing research efforts at NMSU in the areas of computational sciences and data analytics, bringing cutting-edge

techniques and methodologies to the students. In turn, students will have the opportunity of contributing to research projects through course projects, capstone projects, and internships in NMSU research labs.

5. Engagement: the degree program will benefit local communities, by bringing challenge problems suggested by local organizations into the classrooms in the form of class projects.

The specific goals from Vision 2020 that will benefit from the PMCDA program are:

- Academics and Graduation:
    - (KPI 1) Enrollment Growth: the program addresses a focus area of growth and high demand; we expect the program to attract an audience that would currently not consider NMSU as a viable academic option – especially professionals in the workplace.
    - (KPI 2) Graduate Enrollment: the program will be a graduate program at the Master's level, thus promoting increase of graduate enrollment.
    - (KPI 8) STEM-H-B Degrees: the areas of Data Analytics and computational data science are at the core of STEM – being grounded in computer science, mathematics, and statistics. Thus, the PMCDA will provide a new avenue for NMSU to generate STEM-H-B degrees.
- Diversity and Internationalization:
    - (KPI 10) Diversity: the design of the program will build on pedagogical practices developed over the years by the Computer Science department to broaden participation of students from traditionally underrepresented groups in computational sciences. This will have a goal of promoting diversity among the PMCDA students.
    - (KPI 11) International Presence: the PMCDA program will integrate topics and perspectives drawn from ongoing research efforts that rely on international collaborations (e.g., Italy, Germany, Japan, Spain). Students in the program will have the opportunity to meet and interact with international researchers.
- Research and Creative Activities:
    - (KPI 15) Publications and Creativity: the PMCDA program will build on a wealth of research activities in the areas of data science, computer science, and data analytics. Students in the program will have the option of participating in research activities (e.g., as part of their capstone experience) and contribute to discoveries resulting in scientific publications.
- Economic Development and Community Engagement
    - (KPI 16) Student Innovation: the computational data analytics domain is ripe for new discoveries, especially in terms of
        - Exploring applications of computational data analytics methods to novel application domains
        - Exploring how computational data analytics techniques can be advanced to meet the challenges of new application domains.

      These innovations will be led by students as part of their graduate coursework and promoted through existing entrepreneurial avenues (e.g., Arrowhead).

- (KPI 18) Community Engagement: the program will include the opportunity for students, as part of their capstone experience to engage in service learning projects, applying data analytics techniques to the aid of community organizations (e.g., analysis of needs of local shelters based on historical data).
- (KPI 20) Career Placement: as discussed earlier, the workforce demand in the field of applied and computational data analytics is booming, The program will develop collaborations with potential employers to facilitate placement of students upon graduation.

3.3 Relationship to Other NMSU Programs

The PMCDA program is novel and does not represent a duplication of any existing program at NMSU. Nevertheless, the program builds on capacities provided by some existing programs:

- **Computer Science:** the Computer Science programs have expanded in recent years their emphasis on computational methods for data analysis and big data. It provides courses and expertise in database management systems (which is a required area for the Bachelor of Arts in Computer Science), high performance computing, data mining, and machine learning.
- **Applied Statistics:** the program has expanded its expertise in data analysis and mathematical modeling.
- **Industrial Engineering:** the program has developed expertise in data analysis, statistical analysis and data driven decision making.
- **Electrical and Computer Engineering:** the program provides extensive expertise in machine learning and applications (e.g., in signal processing).
- **Mathematical Sciences:** the program provides expertise in basic fundamental mathematical knowledge, mathematical modeling and analysis.

The proposed program is integrative, as it builds on expertise and capacity available across multiple departments in multiple NMSU colleges.

3.4 Relationship to Programs Offered at Other New Mexico Universities

The proposed professional degree program does not represent a duplication of any degree existing degree program offered within New Mexico. The University of New Mexico offers a data analytics concentration within the MS in Information Systems, with emphasis on business applications and data mining/management. Coverage of machine learning and computational methods is very limited. New Mexico Tech does not appear to offer any comparable program.

3.5 Relevant Degree Programs at Peer Institutions

- University of Arizona is currently developing an undergraduate certificate in Data Analytics and Visualization; no comparable graduate programs are offered
- Colorado State University: does not offer any comparable degree program
- University of Idaho: does not offer a comparable degree program
- Iowa State University: offers a Master in Business Analytics, with a greater emphasis on business applications, limited computational aspects, and exclusive focus on big data and data mining (and not machine learning)
- Kansas State University: the only relevant program is a graduate certificate in Data Analytics (15 credits)

- Montana State University: does not offer a comparable program
- University of Nevada-Reno: offers a PhD program in Statistics and Data Science, with a greater emphasis on theoretical and research aspects.
- University of Nebraska-Lincoln: offers a Business Analytics certificate, focused on business professionals and with limited computational coverage.
- University of North Texas: offers a MS degree in Business Analytics, requiring a background in business management and accounting and with emphasis on business applications and data management.
- Oklahoma State University: offers a MS degree in Business Analytics with a good coverage of data mining and programming. It has a strong emphasis on business applications and no coverage of machine learning methods.
- Oregon State University: offers a comprehensive MS degree in Data Analytics, more extensive and advanced with what proposed by NMSU (45 credits)
- University of Texas at El Paso: offers a Graduate Certificate in Data Analytics, with an exclusive emphasis on the mathematical and statistical aspects.
- Texas Tech University: offers a Master's degree in Data Science, but the catalog does not provide a description of the structure of such degree program
- Utah State University: does not offer a comparable degree program
- Washington State University: offers a comprehensive Bachelor degree in Data Science.
- University of Wyoming: offers a Big Data concentration within the Computer Science undergraduate program (focused exclusively on Computer Science majors).

4 Clientele

4.1 Student Characteristics

4.1.1 Students to be Served by the Program

The program will serve a mixed audience composed of

- Traditional on-campus students dedicated to pursuing the degree in a full-time manner
- Professionals in the work-place, pursuing the degree part-time and adapted to the needs of their employers.

We anticipate that in the long run most of our students will be professionals who are looking to add more data analytic tools to their workplace and/or who are seeking advancement or transition to a new functional area. We expect to have both full- and part-time students in the program.

4.1.2 Basic Entry Requirements

All applications for admission in the program will be reviewed by the Director of the program. After an initial screening, the recruitment committee will review applications. Applicants must meet the minimum requirements for admission to the Graduate School (e.g., have completed an undergraduate degree at an accredited institution), and they must have at least one undergraduate course in basic statistics (e.g., equivalent to STAT 251) and a course on linear algebra (e.g., equivalent to MATH 280). Each applicant will be required to provide a statement of interest and three reference letters.

4.1.3 Equitable Representation

We will market the program broadly to help ensure diverse cohorts of students. Given the diversity of students in the NMSU undergraduate student population, we will encourage students NMSU

students to consider PMCDA as an option. We will also work with relevant undergraduate programs (e.g., Computer Science, Mathematical Sciences, Electrical and Computer Engineering, Industrial Engineering) to develop Masters Accelerated Programs that feed into the PMCDA program. We will make additional efforts to market the program in the private sector and in national labs, military installations and other federal and state facilities.

The Department of Computer Science has a long-standing commitment to serve a very diverse student population. NMSU-CS has launched and supports a wide range of initiatives to promote recruitment, training, and retention of students from traditionally under-represented backgrounds. These projects include outreach programs (e.g., summer camps for middle-school and high-school students), training and motivational events (e.g., a year-around set of activities for cohorts of high-school women), and solid links with local high schools and community colleges, and collaborations with Hispanic-Serving Institutions across the nation (e.g., NMSU is one of the leading institutions in the Computing Alliance of Hispanic Serving Institutions).

It is important to underline the importance of the development of this type of program in a region like New Mexico. The field of computing is still witnessing a severe under-representation of women and of students from traditionally under-represented ethnic groups. In particular, less than 30% of computing Master's degrees are of women; only 1.8% of Master's degrees in computing disciplines are awarded to Hispanic students and only 0.2% are awarded to Native American students [11]. The diversity in the population of New Mexico offers an untapped pool of talent on which to build a successful and strong program, laying the foundations for bringing New Mexico to the forefront of training in computational data analytics.

Our program will attempt to address the problem of participation of students from underrepresented backgrounds in graduate studies. We propose to tackle this issue by investing special effort in targeting these students during the recruitment activities. First of all, a significant percentage of our own undergraduate students (over 50%) belong to minority groups. Within these segments of our population there are exceptionally talented students. NMSU is also geographically well-positioned to adequately serve a population with a significant presence of Hispanic and Native Americans. The departments involved have also strong ties with a number of universities and colleges with a predominantly minority population, who will provide another potential pool of talented applicants; we will actively advertise our program at these institutions (during our regular visits). NMSU is part of an Alliance to promote training of Hispanic students in Computing disciplines; this partnership provides resources and a wide network to advertise the program and recruit excellent students from the other partner institutions. Recruitment activities will include visits to other minority-serving colleges and universities, advertisements, presentations, and wide promotion on social media.

4.2 Projected Enrollment

The following table provides a conservative estimate of projected enrollment in the program for the first five years of activity. The figures have been estimated using an internal study of interest in data analytics between existing graduate students, observations of enrollment trends in other data analytics programs. The program will also target a number of part-time students drawn from local industries and research facilities in the area.

Year	**Full-Time**			**Part-Time**			**Completion Rates**
	New	Return	Total	New	Return	Total	
Year 1	5	0	5	0	0	0	0
Year 2	5	5	10	4	0	4	4

Year 3	10	5	15	5	4	9	5
Year 4	10	10	20	5	9	14	14
Year 5	10	10	20	10	10	20	15
Year 6	10	10	20	10	10	20	15

We aim at attracting a sustainable enrollment of about 40 students, with an annual new cohort of about 20 students. We expect to see applications from individuals currently working within a wide variety of economic sectors and industries in the region, due to the interest in data science and data-driven decision making across many economic sectors.

5 Institutional Readiness

5.1 Faculty Resources

The institution has already an adequate pool of qualified researchers and educators to meet the needs of the proposed degree program. The following is a description of the faculty members and their role in the program:

Name	**Department**	**Areas of Expertise**
Charlotte Gard	Applied Statistics	Statistical Analysis, Data Mining, Risk prediction modeling, Health applications
	Courses in the program:	A ST 505, A ST 555
	Rank:	Assistant Professor (FTE: 1.0)
William Gould	Applied Statistics	Sampling and estimation of fisheries and wildlife populations, ecological and biological modeling
	Courses in the program:	A ST 505, A ST 550
	Rank:	Associate Dean
Christopher Sroka	Applied Statistics	Models for count and categorical data, Hierarchical Bayesian models, Simulation
	Courses in the program:	A ST 550
	Rank:	Assistant Professor
David Daniel	Applied Statistics	R Programming
	Courses in the program:	A ST 507, A ST 515
	Rank:	Associate Professor
Enrico Pontelli	Computer Science	High Performance Computing
	Courses in the program:	CS 458, CS 453,
	Rank:	Professor and Dean
Son Cao Tran	Computer Science	Reasoning, Knowledge representation, knowledge-bases
	Courses in the program:	CS 502, CS 506
	Rank:	Professor and Department Head
Huiping Cao	Computer Science	Databases, Data Mining, Big Data
	Courses in the program:	CS 502, CS 508, CS 519
	Rank:	Associate Professor
Zachary Toups	Computer Science	Visualization, Human-Computer interaction
	Courses in the program:	CS 506, CS 454
	Rank:	Assistant Professor

Parth Nagarkar	Computer Science	Databases, Big Data
	Courses in the program:	CS 582, CS 502, CS 508
	Rank:	Assistant Professor
Inna Pivkina	Computer Science	Machine Learning
	Courses in the program:	CS 508, CS 519
	Rank:	Associate Professor
Laura Boucheron	Electrical and Computer Engineering	Machine Learning, Signal Processing
	Courses in the program:	EE 596
	Rank:	Assistant Professor
Phillip De Leon	Electrical and Computer Engineering	Signal Processing, Machine Learning
	Courses in the program:	EE 565
	Rank:	Professor
Talayeh Razzaghi	Industrial Engineering	Data Mining, Machine Learning
	Courses in the program:	IE 460
	Rank:	Assistant Professor
Alla Kammerdiner	Industrial Engineering	Application of data mining to health care, bio-medicine and human factors
	Courses in the program:	IE 567
	Rank:	Assistant Professor
Jameson Cahill	Mathematical Sciences	Applied mathematics, Analysis
	Courses in the program:	
	Rank:	Assistant Professor
Joe Lakey	Mathematical Sciences	Harmonic analysis, discrete dynamic systems
	Courses in the program:	
	Rank:	Professor and Department Head
Tonghui Wang	Mathematical Sciences	Bootstrap methods, multivariate analysis, and linear models in statistics
	Courses in the program:	STAT 470, STAT 480, STAT 535
	Rank:	Professor
Jennifer Kreie	Accounting and Information Systems	
	Courses in the program:	BCIS 485, BCIS 475
	Rank:	Associate Professor
Carlo Mora-Monge	Accounting and Information Systems	
	Courses in the programs:	BCIS 561, BCIS 566
	Rank:	Professor and Department Head

Although *big data*, *data science* and *data analytics* have exploded most rapidly in the business, marketing and healthcare sectors, data science is a pervasive discipline, and therefore it is crucial that the proposed program is interdisciplinary in scope. It is important that its delivery and administration is shared among schools and includes faculty with relevant, applied experience who

represent the different "data using" organization types that the degree serves. Current leaders will also be included as guest lecturers or speakers.

5.2 Library and Curricular Resources

The library provides adequate support for the initial needs of this program. The library provides access to the Digital Library of the Association for Computing Machinery (ACM), which supports the most reputed conferences and journals in Computer Science.

The Department of Computer Science, under the auspices of the NSF CREST grant, has hired a full-time program coordinator who will help with the task of publicizing the program and coordinating the recruitment efforts.

5.3 Physical Facilities

The proposed program does not require any dedicated physical facilities. Nevertheless,

- The general computer laboratory (SH 118) and the graduate laboratory (SH 169) in the Computer Science Department are available to provide computational facilities for students in the program.
- The Business Complex Computer Lab houses 62 Windows-based computers. A variety of software is available including Microsoft Office 2016, R, RStudio, SAS 9.4, MATLAB, Tableau, Desktop GIS, TSP (econometrics software), MiKTeX, and Oracle SQL Developer and SQL Developer Data Modeler. The lab is equipped with two scanners and a webcam and microphone.

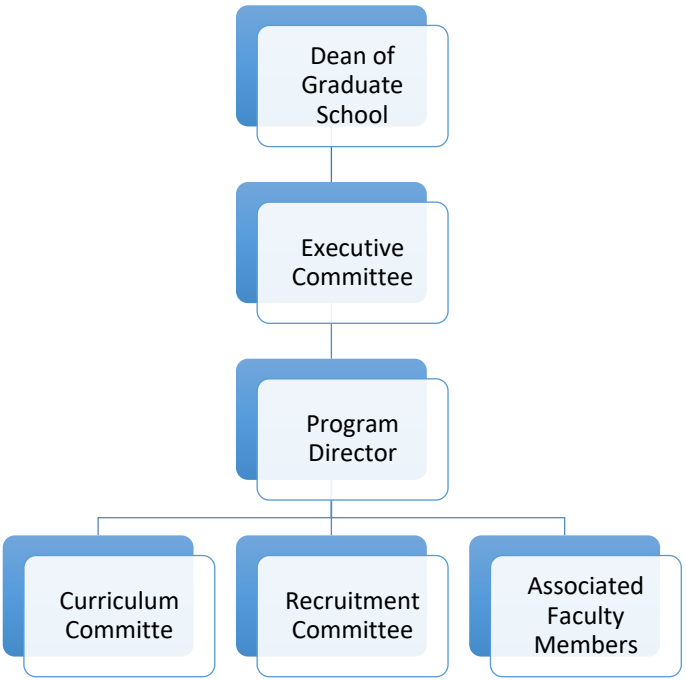
5.4 Equipment and Technology Resources

The courses in the program will make exclusive use of public domain and freely available software infrastructures. An existing National Science Foundation award will also provide a computing server that will be dedicated to the needs of this academic degree program – by providing storage space and installation of software (e.g., R, Python, database systems, machine learning frameworks) for the needs of the students in the program. Access to the server and its resources will be restricted to students admitted in the program and registered for courses in the program.

5.5 Administrative Structure

The proposed degree program will be an inherently interdisciplinary program. The academic governance will be by an Executive Committee consisting of the Deans (or their designated representatives) of the Colleges of Arts & Sciences, Engineering, and Business. The Executive Committee is designed to provide overall interdisciplinary program and operational review. It will determine and review the Program's policies and procedures.

The program will exist independent of any department, but will draw on faculty from participating departments. The program will be led by a Program Director, who will be a full time (9 month) tenured (or tenure track) faculty member. The Director will report to the Executive Committee on curriculum matters, manage the recruitment and admissions process, develop and maintain relationships with other relevant degree programs, maintain connection with industry and government entities to facilitate internship opportunities, teach in the program, chair the program's curriculum committee, implement and maintain courses, and report annually (or more frequently if necessary) to the Executive Committee. Because it is an interdisciplinary program the nexus of both operational and organizational structures will be with the Graduate School. The following chart outlines the governance structure.



6 Summary of Costs and Benefits

The proposed program should incur minimal new costs as almost all elements of the proposed program are already in place. The courses required to complete the program are already listed in the graduate course catalog and are being offered by various departments on a regular basis. We anticipate that graduate student assistance will be needed with the course instruction. We anticipate that we will need **two** graduate assistants to help with the instruction of the courses. Initially, this support for the assistants will come from the external grants that are already in place. The $5 million CREST (Center for intelligent technologies for smartgrids) grant that started in 2014 and will run through at least 2020 provides support for about 10 graduate assistants per semester. Some of these students will help with the instruction of the courses. We anticipate that the CREST grant will be funded for another five year period. However, in the case that the CREST grant is not renewed, NMSU will provide the support for the two additional GAs needed to support instruction of the courses required for the program (commitment from the College of Arts & Sciences).

6.1 Projected Costs

We do not anticipate any substantial costs to start-up the program. Below we provide a brief analysis for each of the required categories.

6.1.1 Additional faculty needed for the program

As stated earlier, the faculty needed for the program is already in place. The idea and planning of a Masters degree in Computational Data Analytics has been going on at NMSU for a while. One new faculty member, Dr. Parth Nagarkar (Computer Science), with expertise in Data Management was hired in 2017, joining an existing cadre of faculty members dedicated to this discipline. We anticipate that at least one additional faculty member will be hired over the next year who will strengthen the program – specifically, bringing expertise in data visualization and presentation. However, the development of the program is not contingent upon these additions.

6.1.2 Additional library resources needed for the program.

The proposed program is a Masters program. Hence the need for additional library resources is not as extensive as it might be for a new doctoral program. The library resources as they exist are adequate for the proposed program. The library has substantial holdings in computer science and related disciplines. Almost all research materials that might be needed for the program are accessible through Web resources or inter-library loans. Furthermore, there is a wealth of materials that is freely available online, and faculty members are used to prioritize public domain materials in their courses.

6.1.3 Additional facilities, equipment and technological resources

We believe that the existing facilities, equipment and technological resources are sufficient for the program. The Computer Science department has a graduate computing lab and significant computing resources to support all additional computing needs of the program. Additional equipment (e.g., one new server, several new workstations) will be added to these facilities in 2018 through existing funding from the National Science Foundation. While this will strengthen the training component of the MS program, the existence and continuation of the degree is not contingent upon the setting up of the lab.

6.1.4 New graduate assistantships needed to support the program

Two new graduate assistantships will be needed to support the program. As indicated earlier, there are significant existing external grants at NMSU that will mitigate the cost of the program for several years initially. We anticipate that we will not need any additional graduate assistantships for at least the first three years. Thereafter, we still anticipate that external grants will fund the two additional assistants needed to support the instruction of the program courses.

6.2 Projected Benefits

6.2.1 Financial Benefit

The following table illustrates the total revenues in terms of tuition and fees conservatively estimated based on an average enrollment of 18 credits per year for full-time students and 12 credits per year for part-time students, for in-state students ($291.3 per credit hour in tuition and fees).

Year	**Full-Time**			**Part-Time**			**Tuition and Fees**
	New	Return	Total	New	Return	Total	
Year 1	5	0	5	0	0	0	$26,217
Year 2	5	5	10	4	0	4	$66,416
Year 3	10	5	15	5	4	9	$110,111
Year 4	10	10	20	5	9	14	$153,806
Year 5	10	10	20	10	10	20	$174,780
Year 6	10	10	20	10	10	20	$174,780

6.2.2 Benefits to the State of New Mexico

Benefits to the NMSU System: The very nature of data analytics cuts across academic disciplines, allowing for diverse interdisciplinary collaborations and expanded potential for discovery. The proposed degree structure allows for future growth, by adding concentration areas to meet the analytics needs of different disciplines. This will enable new collaborations not only within the institution, but also business, industry, and government. The proposed degree will help create a community of faculty from Computer Science, Engineering, Applied Statistics, Mathematics,

Business, Education, and other disciplines dedicated to big data and analytics and interested in graduate level Data Analytics and Analytics education. At present there is no locus for interaction related to graduate level data analytics education within the NMSU system.

Research, Teaching, and University Benefits: The proposed PMCDA program provides a number of positive benefits to NMSU. For example, this program will lay the academic foundations and the institutional commitment necessary to recruit highly skilled data scientists for university oriented business development. Bringing data analytics to the NMSU campus would provide a significant competitive advantage for New Mexico's business and industry, state agencies, national laboratories, and universities, and thus enhance the state's economic development.
To achieve these benefits, we will develop an online clearinghouse for matching research and practical projects with students in the PMCDA program – to be used as part internships and culminating experiences for the students in the program. The aim is to bring computational data analytics capabilities to campus wide activities.

Educational Benefits to New Mexico's Students: An essential component of the PMCDA program is a sequence of courses aimed at providing students with a specialization/emphasis. This offers in-depth analyses and training on data analytic techniques, issues, and problems students will face within a given concentration area. Students will have the opportunity, within such concentration, to gain hands-on experience with large and complex data sets and use cutting-edge computational analytics techniques. During such courses, the students will benefit from the direct mentoring from faculty, as well as insight from industry partners.
The proposed degree, with its professional structure, will prepare students with different backgrounds to become productive data scientists in New Mexico and regionally connected laboratories, government, and industries.
The PMCDA program reflects, in its interdisciplinary structure, the diversity of the emerging field of data science by offering emphases in several different disciplines in advanced data science involving big data. The degree has a professional structure, with an estimated time to degree of four (full-time) to six (part-time) semesters, with an approximate total cost of $xxxxx.

7 Assessment of Operations and Impact

The program will conduct regular assessments of learning outcomes and report its findings on an annual basis to the executive committee, the Dean of the Graduate School and the Provost. We will collect reliable direct and indirect evidence of student learning, and ensure that assessment closes the loop, leading to improvements in the curriculum, teaching methodologies and learning outcomes.
The proposed degree program will align its curriculum so that it will be consistent with the learning objectives highlighted in the earlier sections of this document. The learning outcomes will be assessed with a focus on coursework learning outcomes, research projects/internship outcomes, data scripting outcomes, testing, conformance to best practices, student portfolios, surveys, class presentations and a variety of evaluations.

7.1 Evaluation and Assessment

The faculty involved in the PMCDA program will follow the assessment standards for graduate education at NMSU, using the guidelines established by the Annual Academic Departmental Assessment of Student Learning. The program faculty will assess the intellectual development and

learning outcomes of students through various formative and summative means at the course level and at the program level. A mixed-method program assessment approach will employ direct and indirect assessment measures and will be used to inform pedagogical, structural, and learning outcome modifications as the program matures.

- **Direct Measures:** The PMCDA program will review all students' deliverables (e.g., course projects, sample of course student work, portfolios, research projects) with rigorous intellectual engagement of students as a primary consideration. Exams/tests, papers, projects, presentations, portfolios, etc. will capture what students actually learn and their mastery of these skills. The executive team will work with course instructors in developing and deploying standard assessment rubrics.
- **Indirect Measures:** During the initial years of operation, the program will use outside summary reports developed by an evaluation team, assembled by the Graduate Dean and coordinated through the NMSU Office of Assessment. The summary reports will assess the alignment of essential and important objectives with the stated objectives of the program. Additionally, the evaluation team summary reports will allow for a review of students' perception of their progress, the rigor of the courses, and the pedagogy that faculty used in comparison to the interdisciplinary and college norms. All PMCDA graduates will complete an exit survey to assess both the processes and the outcomes of their courses and their interactions with faculty and fellow students. Other indirect measures such as surveys, interviews, and aggregate reports on retention, graduation, and placement will be utilized.

The Director of the PMCDA program will lead appropriate process assessments to fine-tune the operation and provide guidance and adjustment to course instructors and curriculum developers. Evaluation and assessment of the program will follow the established processes for Outcome Assessment in place at NMSU.

Summative evaluations will include surveys of continuing and graduating students that will be conducted each year, along with faculty evaluations and input from the previously mentioned evaluation committee. This committee will overview curriculum and the admissions process. A focus group approach will be conducted each year with graduating students and one year alumni. The focus groups will determine student perspectives on the curriculum and instruction. The alumni will provide feedback on job readiness and workforce experience. Employment placement of students and their rate of compensation will be key indicators of program quality. Employers will be surveyed periodically to ascertain their views of the programs and to aid in determining what skills should be emphasized (or not) in the curriculum.

Formative assessment will include student coursework, drawn from Canvas completed submissions, and portfolio reviews. Faculty will meet once a year to review student progress and quality, and provide progress reviews/feedback to students and assigned advisors. In addition, the program will undergo the yearly Outcome Assessment review and will be integrated in the 5-year Program Review process (e.g., as part of one of the lead departments).

Consistent with the program objectives described in earlier sections of this document, the program goals (which include core competencies and technical skills) include:

- Identify and describe data analysis problems

- Gather and organize homogeneous and heterogeneous data from public and private data sources, including web mining and database searches
- Prepare datasets for analysis, through integration, harmonization, scaling, cleaning, converting and filtering data
- Analyze data for patterns, trends
- Determine classification and predictive models based on data
- Develop summaries, visualization and other descriptive models
- Apply the most appropriate analytics instruments to data sets
- Use data analytics techniques to creatively solve problems

Different courses will introduce the most appropriate technologies (e.g., Weka), but all students will gain expertise in at least one operating system and two programming environments (Python and R).

8 Other

8.1 Accreditation

At present there is no recognized accrediting body for programs of this sort. Data analytics programs are quite new; they come in many different varieties, and there is no consensus as yet regarding what these programs should look like.

9 Appendices

9.1 Appendix A: Courses

A ST 505. Statistical Inference I [4 Credits]
A qualitative introduction to the concepts and methods of statistical inference. Sampling, frequency distributions (z, t, x2, F), estimation, and testing. One-way analysis of variance. Simple linear regression.

A ST 507. Advanced Regression [3 Credits]
Examination of multiple regression; residual analysis, collinearity, variable selection, weighted least squares, polynomial models, and nonlinear regression: linearizable and intrinsically nonlinear models.

A ST 515. Statistical Analysis with R [3 Credits]
Introduction to R data types, basic calculations and programming, data input and manipulation, one and two sample tests, ANOVA, regression, diagnostics, graphics, probability distributions, and basic simulations in the R software environment.

A ST 550. Predictive Analytics [3 Credits]
Data analytics techniques, data pre-processing, model tuning and cross-validation, linear regression, non-linear regression, regression trees, classification models.

A ST 555. Applied Multivariate Analysis [3 Credits]
Multivariate analysis of linear statistical models, including MANOVA and repeated measures. Analysis of correlation and covariance structures, including principal components, factor analysis, and canonical correlation. Classification and discrimination techniques.

ASTR 630. Methods of Statistical Analysis for Modern Astronomy [3 Credits]
Graduate class for students interested in applying statistical techniques to modern astrophysical data sets. Topics include a review of probability and probability distribution functions,

implications of techniques for statistical inference, regression and multivariate analysis, data smoothing, data mining, survival analysis and time domain analysis. Applications to real astronomical data sets will be emphasized with all topics.

BCIS 575. Database Management Systems [**3 Credits]**
Design, development, and use of database management systems in the business environment.
BCIS 585. Enterprise Resource Planning [**3 Credits]**
This course covers concepts in enterprise resource planning (ERP). Topics include how ERP integrates business processes across functional areas--such as the procurement process and the sales order process--and how businesses use ERP information systems in day-to-day operations as well as for performance monitoring. SAP R/3 software will be used in several hands-on examples of ERP software as a real-world example of an ERP system.
BCIS 561. Business Analytics I [**3 Credits]**
This course provides an understanding of how organizations can utilize technology to successfully collect, organize, manipulate, use, and present data. The course blends the use of current technology with the managerial practices involving business analytics. The emphasis of the course will be on data management practices and the production of descriptive analytics.
BCIS 566. Business Analytics II [**3 Credits]**
This course provides an understanding of how organizations can build and test predictive models, utilizing business-related data to estimate model parameters. The emphasis of the course will be on utilizing data management systems to produce useful predictive analytics.

BIOL 566. Advanced Bioinformatics and NCBI Database [**3 Credits]**
The course discusses how to use NCBI database and bioinformatic tools for research with genomics approaches. The topics include nucleotide and protein sequence analysis, similarity search with blast algorithms, gene/genome annotation, protein structure analysis, gene expression analysis, and metagenomic study.

C S 453. Python Programming I [**3 Credits]**
This course is an introduction to programming in the Python language, covering fundamental scripts, data types and variables, functions, and simple object creation and usage. The focus will be on preparing students to use Python in their own areas.
C S 454. Python Programming II [**3 Credits]**
This course covers advanced Python programming, including classes, objects, and inheritance, embedded programming in domain applications, database interaction, and advanced data and text processing. The focus will be on preparing students to use Python in their own areas.
C S 458. R Programming I [**3 Credits]**
This course is an introduction to data processing in the R language, covering fundamental script configuration, data types and data collections, R control structures, and basic creation of graphs and data visualizations. This course will not focus on the statistical capabilities of R, though some basic statistical computations will be used.
C S 459. R Programming II [**3 Credits]**
This course covers advanced R programming, including advanced data collection processing, advanced data visualizations, object oriented features of R, and file processing. It is recommended that students have one statistics course before taking this course.
C S 502. Database Management Systems I [3 Credits]

Database design and implementation; models of database management systems; privacy, security, protection, recovery; Students are expected to have solid knowledge of data structures and discrete mathematics.

C S 506. Computer Graphics I [3 Credits]

Languages, programming, devices, and data structures for representation and interactive display of complex objects.

C S 508. Introduction to Data Mining [3 Credits]

Techniques for exploring large data sets and discovering patterns in them. Data mining concepts, metrics to measure its effectiveness. Methods in classification, clustering, frequent pattern analysis. Selected topics from current advances in data mining.

C S 512. Computer Systems Modeling and Simulation [3 Credits]

Basic concepts of modeling computer systems: continuous and discrete time models, states and transition, probabilistic models. Structures of simulation programs, time driven and event driven simulation, simulation on captured and synthetic traces, generation of random variables, queuing models, Markov chains, random walks, Poisson, Markov, renewal branching and Brownian motion processes, model validation and data analysis.

C S 519. Applied Machine Learning I [3 Credits]

An introductory course on practical machine learning. An overview of concepts for both unsupervised and supervised learning. Topics include classification, regression, clustering, and dimension reduction. Classical methods and algorithms such as linear regression, neural networks, support vector machines, and ensemble approaches. Recent techniques such as deep learning. Focused on applying of machine learning techniques in application domains. More work might be required for graduate students.

COMM 550. Seminar in Communication Technologies [3 Credits]

Seminar on design, usage, and social impact of electronic mail, communication through computer networks, and new technologies of organizational communication such as group decision support systems (GDSS). Each student will study an actual application of a major communication technology in an organization.

E E 565. Pattern Recognition and Machine Learning [3 Credits]

Statistical pattern classification, supervised and unsupervised learning, feature selection and extraction, clustering, image classification and syntactical pattern recognition.

E E 596. Digital Image Processing [3 Credits]

Two-dimensional transform theory, color images, image enhancement, restoration, registration, segmentation, compression and understanding.

ENGL 543. Multimedia Theory and Production [3 Credits]

Issues, theories, and production practices underlying design of multimedia, including rhetorical choices, aesthetic approaches, usability concerns, and diverse academic and popular discourses contributing to continued development of digital texts.

ICT 458. Database Design and Applications [3 Credits]

MySQL and PHP. Data conversion using PHP, mysql and Python. Methods of transferring data from electronic boards and data feeds, into databases. Use of SQL in java programming. Remote

programming of computers for running database systems in a mixed OS environment. Generation of web pages directly from Database queries.

ICT 460. Web Technologies and Multimedia [**3 Credits]**

Addresses the latest multimedia technology advances and how they apply to the information and communication technology fields.

I E 460. Evaluation of Engineering Data [**3 Credits]**

Analysis of engineering systems possessing variability, employing regression, analysis of variance, distribution theory, and experimental design methods.

I E 515. Stochastic Processes Modeling [**3 Credits]**

Introduction to the use of stochastic processes in the modeling of physical and natural systems. Use of generating functions, conditional probability and expectation, Poisson processes, random walk models, Markov chains, branching processes, Markov processes, and queuing processes in an applied setting.

I E 522. Queuing Systems [3 Credits]

Elements and classification of queuing systems, single server models, multi-server models, cost analysis and applications.

I E 537. Large Scale Systems Engineering [3 Credits]

Systems engineering approaches to large-scale complex technological and societal problems. Concepts of interaction and structural graphs, matrices, delta, and Gantt charts. The hall matrix approach, structural concepts, reachability matrices, and cross impact-analysis, modeling and decision making.

I E 545. Characterizing Time-Dependent Engineering Data [3 Credits]

Theory and techniques employed in the characterization of stochastic processes commonly found in engineering applications. Distribution models include exponential, gamma, Weibull, and extreme value. Design and analysis of experiments involving complete and censored data and elevated stress. Analytical techniques include parametric, nonparametric, and graphical approaches with emphasis on modern computer tools. Exact and approximate maximum-likelihood techniques are stressed.

I E 567. Discrete-Event Simulation Modeling [**3 Credits]**

Basic modeling concepts, organizations of simulations, input data analysis, random variate generation, simulation design and analysis, model validation, output analysis, and management of simulations.

MATH 518. Fourier Series and Boundary Value Problems [**3 Credits]**

Fourier series and methods of solution of the boundary value problems of applied mathematics.

STAT 470. Probability: Theory and Applications [**3 Credits]**

Basic probability distributions including binomial, normal; random variables, expectation; laws of large numbers; central limit theorem.

STAT 480. Statistics: Theory and Applications [**3 Credits]**

Point and interval estimation; sufficiency; hypothesis testing; regression; analysis of variance; chi-square tests.

STAT 535. Elementary Stochastic Processes [**3 Credits]**

Markov chains, Poisson processes, Brownian motion, branching processes, and queuing processes, with applications to the physical, biological, and social sciences.

9.2 Representative Programs in Other Universities

The following is a list of popular Master's degree programs according to several web site and journal rankings:

American University: The program is part of the Kogod School of Business. The program is a professional degree with emphasis on the business analytics aspects and with no computational competencies (beyond management of database systems). The degree requires 33 credits, which include 12 credits of elective courses in one of three areas of specialization: Policy, Finance, Marketing.

Syracuse University: the institution offers two degrees. The first one is a Master of Science in Business Analytics. The degree is a professional online degree, with an emphasis on applications in accounting, marketing, and business problems decision making. The second degree is a Master in Applied Data Science. This is also a professional degree program, with an emphasis on the application of tools and techniques to provide data-driven solutions. The program is composed of 36 credits, does not appear to offer any emphasis on the computational aspects and on the mathematical underpinnings.

Southern Methodist University: this institution offers a Master of Data Science degree program; the program has analogies to what we propose at NMSU. The program provides a rigorous introduction to the statistical and mathematical foundations, applied data mining, and a capstone course. Differently from what we propose, the program does not address computational aspects (e.g., programming) – a coding course is available as an elective.

University of California-Berkeley: The Master in Information and Data Science is composed of 27 credits and is the closest in structure to what we propose for NMSU. It includes foundational courses in both programming (Python) and statistics; differently from our approach, it emphasizes machine learning over data mining.

George Washington University: this institution offers a Master's degree in Management of Health Informatics and Analytics; the emphasis of the program is on management of health records with application to public health, epidemiology and health marketing. The program does not provide emphasis on computing or machine learning.

Maryville University: The Master's degree in Business Data Analytics requires 30 credits, organized in four areas: foundational courses (programming, databases), descriptive analytics (warehousing and analytics), predictive analytics (data mining and visualization), and prescriptive analytics (forecasting and capstone). The emphasis is only modestly on computing and there is a greater emphasis on data storage and management, and not on machine learning.

Arizona State University: the Master of Science in Business Analytics is focused on data mining, marketing, and business analytics; the program does not offer foundations of programming and statistics and no coverage of machine learning.

Villanova University: the Master of Science in Analytics consists of 36 credits; it includes a capstone course and places emphasis on both introductory programming (in R), data mining, business applications and use of data mining in enterprise systems. The program has limited computational coverage and no coverage of machine learning. The program does provide some foundational courses on multivariate analysis and statistical modeling.

Southern New Hampshire University: the Master in Data Analytics is composed of 36 credits, with a strong emphasis on data management, decision methods, risk assessment and project management. The degree does not provide statistical methods, machine learning or any level of programming.

Purdue University: the Master of Science in Business Analytics and Information Management requires 36 credits, organized in four areas: core courses (mostly focused on IT management, data management, data mining based on specific tools), communication and persuasion (1 course), functional area courses (mostly in accounting, economics, and management), electives (e.g., web analytics, R, spreadsheets). There is limited programming content, strong emphasis on business applications, no coverage of statistical modeling and machine learning.

Utica College: the Master of Science in Data Science is the closest degree program to what we propose. Composed of 30 credits, it includes a set of core courses (statistical methods, machine learning, data mining), electives organized by tracks (e.g., business, security, social sciences) and a capstone experience. Differently from what we propose, there is no programming content.

Merrimack College: the Master of Science in Data Science is a comprehensive degree program covering modeling, machine learning, data mining, visualization and social implications of data science. The degree includes a multi-source analysis capstone component; there is no programming component.

Saint Joseph's University: the Business Intelligence and Analytics MS requires 30 credits, with a good coverage of data mining and programming (in both Python and R), along with applications in business analytics and supply-chain management. The program does not cover statistical modeling or machine learning.'

Penn State World Campus: the Master of Professional Studies in Data Analytics is composed of 30 credits, composed of a core (data mining and applied statistics), prescribed courses (databases and data-driven decision making), a selection of electives (including a variety of applied statistics and data mining and visualization courses), and a capstone course. The program does not appear to cover programming or machine learning.

10 References

[1] *Master's in Data Science*, http://www.mastersindatascience.org/schools/new-mexico/, 2017.

[2] BURNING GLASS TECHNOLOGIES, *The Quant Crunch: How The Demand For Data Science Skills Is Disrupting The Job Market*, Burning Glass Technologies, Business-Higher Education Forum (BHEF), and IBM,, https://public.dhe.ibm.com/common/ssi/ecm/im/en/iml14576usen/IML14576USEN.PDF, 2017.

[3] L. BURTCH, *Salaries of Data Scientists*, *The Burtch Works Study*, Burtch Works Executive Recruiting, 2016.

[4] BURTCH WORKS, *2017 Data Science Hiring Market*, http://www.burtchworks.com/2017/04/24/key-trends-shaping-2017-data-science-hiring/, 2017.

[5] L. CAO, *Data Science: A Comprehensive Overview. ACM Computing Surveys*, ACM Computing Surveys, 50 (2017).

[6] INTERNATIONAL DATA CORPORATION, *Double-Digit Growth Forecast for the Worldwide Big Data and Business Analytics Market Through 2020 Led by Banking and Manufacturing Investments, According to IDC* https://www.idc.com/getdoc.jsp?containerId=prUS41826116, 2015.

[7] J. KING and R. MAGOULAS, *2016 Data Science Salary Survey*, O'Reilly, http://www.oreilly.com/data/free/files/2016-data-science-salary-survey.pdf, 2016.

[8] J. MANYIKA, M. CHUI, B. BROWN, J. BUGHIN, R. DOBBS, C. ROXBURGH and A. H. BYERS, *Big data: The next frontier for innovation, competition, and productivity*, McKinsey Global Institute, 2011.

[9] MCKINSEY GLOBAL INSTITUTE, *The Age of Analytics: Competing in a Data-driven World*, http://www.mckinsey.com/business-functions/mckinsey-analytics/our-insights/the-age-of-analytics-competing-in-a-data-driven-world, 2016.

[10] NEW MEXICO STATE UNIVERSITY, *Vision 2020 Strategic Plan*, https://plan.nmsu.edu/, 2015.

[11] S. ZWEBEN and B. BIZOT, *2016 Taulbee Survey*, COMPUTING RESEARCH NEWS, 29 (2017).