Centro Federal de Educação Tecnológica Celso Suckow da Fonseca - CEFET/RJ

Programa de Pós-Graduação em Ciência da Computação

PROCESSO SELETIVO 2018.1

PROVA DE LÍNGUA INGLESA

INSTRUÇÕES GERAIS AOS CANDIDATOS

- O tempo total para realização das provas é de **1 hora** ***(1h)***.
- Ao término da prova o candidato deverá devolver o cartão resposta.
- É imprescindível verificar no cartão resposta o número de inscrição do candidato no espaço reservado para tal.

  A IDENTIFICAÇÃO DOS CANDIDATOS EM TODAS AS PÁGINAS DEVERÁ SER FEITA **APENAS** PELO NÚMERO DE INSCRIÇÃO.
- As respostas deverão ser transpostas para o cartão resposta com caneta de tinta azul ou preta. Não serão consideradas as respostas que não estiverem transcritas no cartão resposta bem como não serão consideradas respostas rasuradas.
- A prova de Língua Inglesa é constituída por 10 questões objetivas.
- Cada questão objetiva tem somente uma resposta correta.
- A prova deve ser feita sem consulta e sem empréstimo de material.
- Verifique se sua prova contém 10 questões, assim como o cartão de respostas.
- **Não** é permitido o uso de calculadora, celular ou qualquer outro aparelho durante a realização da prova. É vedado o empréstimo de qualquer material entre os candidatos.

Boa Prova !

Centro Federal de Educação Tecnológica Celso Suckow da Fonseca - CEFET/RJ

Programa de Pós-Graduação em Ciência da Computação

PROCESSO SELETIVO 2018.1

PROVA DE LÍNGUA INGLESA

TEXT 1

Introduction: Becoming a Unicorn

"Data Science" is a very popular term these days, and it gets applied to so many things that its meaning has become very vague. So I'd like to start this book by giving you the definition that I use. I've found that this one gets right to the heart of what sets it apart from other disciplines. Here goes:

> Data Science means doing analytics work that, for one reason or another, requires a substantial amount of software engineering skills.

Sometimes, the final deliverable is the kind of thing a statistician or business analyst might provide, but achieving that goal demands software skills that your typical analyst simply does not have. For example, a dataset might be so large that you need to use distributed computing to analyze it or so convoluted in its format that many lines of code are required to parse it. In many cases, data scientists also have to write big chunks of production software that implement their analytics ideas in real time. In practice, there are usually other differences as well. For example, data scientists usually have to extract features from raw data, which means that they tackle very open-ended problems such as how to quantify the "spamminess" of an e-mail.

It is very hard to find people who can construct good statistical models, hack quality software, and relate this all in a meaningful way to business problems. It is a lot of hats to wear! These individuals are so rare that recruiters often call them "unicorns".

The message of this book is that it is not only possible but also relatively straightforward to become a "unicorn". It is just a question of acquiring the particular balance of skills required. Very few educational programs teach all of those skills, which is why unicorns are rare, but that is mostly a historical accident. It is perfectly reasonable for a single person to have the whole palette of abilities, provided they are willing to ignore the traditional boundaries between different disciplines.

This book aims to teach you everything you will need to know to be a competent data scientist. My guess is that you are either a computer programmer looking to learn about analytics or more of a mathematician trying to bone up on their coding. You might also be a businessperson who needs the technical skills to answer your business questions or simply an interested layman. Whoever you are though, this book will teach you the concepts you need.

This book is not comprehensive. Data science is too big an area for any person or book to cover all of it. Besides, the field is changing so fast that any "comprehensive" book would be out-of-date before it came off the press. Instead, I have aimed for two goals. First, I want to give a solid grounding in the big picture of what data science is, how to go about doing it, and the foundational concepts that will stand the rest of time. Second, I want to give a "complete" skill set, in the sense that you have the nuts-and-bolts knowledge to go

out and do data science work (you can code in Python, you know the libraries to use, most of the big machine learning models, etc.), even if particular projects or companies might require that you pick up a new skill set from somewhere else.

CADY, Field. The Data Science Handbook. NJ: John Wiley & Sons, 2017. P 1-2.

1. Why is the title of TEXT 1 "Becoming a Unicorn"?

A. Because unicorns are mythological creatures which represent strength, a quality needed by data scientists.

B. As data scientists need to combine many abilities and it is not an easy task, these professionals are as difficult to be found as unicorns.

C. Because unicorns combine features of horses and goats as data scientists need to combine abilities from different areas.

D. As data scientists are able to solve all open-ended problems, these professionals sometimes can be seen as having special powers as unicorns.

2. Qual das opções abaixo contém um falso cognato?

A. Typical ("...that goal demands software skills that your typical analyst simply doesn't have.")

B. Implement ("...have to write big chunks of production software that implement their analytics ideas in real time.")

C. Comprehensive ("This book is not comprehensive.")

D. Solid ("First, I want to give a solid grounding in the big picture of what data science is.")

3. According to TEXT 1:

A. The difference between a statistician analyst and a data scientist is that the former has great software skills.

B. It's pretty complex to become a data scientist once the professional needs to acquire different skills and courses can be pricey.

C. The Data Science Handbook intends to be the most complete guide for data scientists and it promises to deal with all the facts and figures of the field.

D. The author of the book wants it to be practical and to show readers the essential aspects of the area.

4. The expression "It is a lot of hats to wear" (3rd paragraph) relates to the fact that:

A. A data scientist is the kind of person who performs different roles.

B. A data scientist is the kind of person who spends time doing one type of business.

C. A data scientist is the kind of person who can easily network.

D. A data scientist is the kind of person who focus on a tool and manages it well.

5. In the sentence, "This book is not comprehensive. Data science is too big an area for any person or book to cover all of it. Besides, the field is changing so fast..." (last paragraph), besides could be substituted for

A. However.

B. Nevertheless.

C. On the other hand.

D. Furthermore.

TEXT 2

Data science is a type of "intelligence science" that aims to transform data into knowledge, intelligence, and wisdom. In this transformation, comprehensive intelligence, or "X-intelligence", is often used to address a complex data science problem, referring to comprehensive and valuable information. X-intelligence can help inform the deeper, more structured and organized comprehension, representation, and problem solving in the underlying complexities and challenges.

Data intelligence highlights the most valuable information and narratives in the formation and solution of business problems or value in the corresponding data. Intelligence hidden in data is discovered by data science through its ability to understand data characteristics and complexities. Apart from the usual focus on complexities in data structures, distribution, quantity, speed, and quality, the focus in data science is on the intelligence hidden in the unknown. For example, in addition to existing protocols for cancer treatment, determining what new and existing treatments fail on which patients might be informed by analyzing healthcare data and diversified external data relevant to cancer patients. The level of data intelligence depends on how much and to what extent a data scientist is able to deeply understand and represent data characteristics and complexities.

Data scientists discover behavior intelligence by looking into the activities, processes, dynamics, and impact of individual and group actors, or the behavior and business quantifiers, owners, and users in the physical world. Such discovery requires they be able to bridge the gap between the data world and the physical world by connecting what happened and what will happen in the problem and discovering behavior insights through behavior informatics. For example, in monitoring online shopping websites, regulators must be able to recognize whether ratings and comments are made by robots, rather than humans; likewise, in social

media, detecting algorithms, or robot-generated comments, in billions of daily transactions is itself a computational challenge. Constructing sequential behavior vector spaces and modeling interactions with other accounts in a given time period and then differentiating abnormal behaviors may be useful for understanding the difference between proactive and subjective human activity and the reactive and patternable behaviors of software robots.

Domain intelligence emerges from relevant domain factors, knowledge, meta-knowledge, and other domain specific resources associated with a problem and its target data. Qualitative and quantitative domain intelligence can help inform and enable a data scientist's deep understanding of domain complexities and their roles in discovering unknown knowledge and actionable insight. For example, to learn highfrequency trading strategies for use with stock data, a strategy modeler must include the "order book" and microstructure of the related "limit market".

Human intelligence plays a central role in complex data science systems through explicit, or direct, involvement of human intuition, imagination, empirical knowledge, belief, intention, expectation, runtime supervision, evaluation, and expertise. It also concerns the implicit, or indirect, involvement of human intelligence in the form of imaginative thinking, emotional intelligence, inspiration, brainstorming, reasoning inputs, and embodied cognition, as in convergent thinking through interaction with fellow humans. For example, as "data-science thinking" is crucial for addressing complex data problems, data scientists must be able to apply subjective factors, qualitative reasoning, and critical imagination.

Network intelligence emerges from both Web intelligence and broad-based networking and connected activities and resources, especially through the Internet of Things, social media, and mobile services. Information and facilities from the networks involved in target business problems can contribute useful information for complex data-science problem solving; a relevant example is crowdsourcing-based open source system development and algorithm design.

Organizational intelligence emerges from the proper understanding, involvement, and modeling of organizational goals, actors, and roles, as well as structures, behaviors, evolution and dynamics, governance, regulation, convention, process, and workflow in data science systems. For example, the cost effectiveness of enterprise analytics and functioning of data science teams rely on organizational intelligence.

Social intelligence emerges from the social complexities discussed earlier. Human social intelligence is embedded in social interactions, group goals and intentions, social cognition, emotional intelligence, consensus construction, and group decision making. Social intelligence is also associated with social network intelligence and collective interactions among social systems, as well as the business rules, law, trust, and reputation for governing social intelligence. Typical artificial social systems include social networks and social media in which data-driven social complexities are understood through social influence modeling, latent relation modeling, and community formation and evolution in online societies.

Environmental intelligence is also hidden in data science problems, as specified in terms of the underlying domain and related organizational, social, human, and network intelligence. Data science systems are open, with interactions between the world of transformed

data and the physical world functioning as the overall data environment. Examples include context-aware analytics involving contextual factors and evolving interactions and changes between data and context, as in infinite-dynamic-relation modeling in social networks.

CAO, L. Data Science: Challenges and Directions. In: Communications of the ACM. August 2017, vol. 60, n. 8, p. 61, 62.

6. The excerpt above is part of a text about data science. A subtitle was given to each specific part. Which subtitle suits this part of the text?

A. X-Intelligence in Data Science
B. Understanding data characteristics and complexities in Data Science
C. The roles of Data Science
D. Hidden Intelligences in Data Science

7. De acordo com o TEXTO 2, considere as afirmativas abaixo:

I A ciência de dados pode ajudar na descoberta de doenças como o câncer.

II A capacidade do cientista de dados em entender profundamente e representar as complexidades dos dados será determinante para o nível de inteligência dos dados.

III O foco da ciência de dados está na inteligência escondida no desconhecido.

Marque a opção que contenha a(s) proposição(ões) correta(s).

A. I
B. II
C. I e III
D. II e III

8. A eficácia dos custos do funcionamento de equipes de cientistas de dados está relacionada à

A. Inteligência comportamental
B. Inteligência organizacional
C. Inteligência social
D. Inteligência ambiental

9. The word regulators in "...in monitoring online shopping websites, regulators must be able to recognize..." (3rd paragraph) is formed by a suffix which makes nouns mean "someone or something that performs an action". The option containing a noun with a suffix that conveys the same meaning as in regulator is:

A. Deeper ("X-intelligence can help inform the **deeper**..." – 1st paragraph)

B. Discover ("Data scientists **discover** behavior intelligence..." – 3rd paragraph)

C. Actor ("...and impact of individual and group **actors**..." – 3rd paragraph)

D. Behaviors ("...and roles, as well as structures, **behaviors**, evolution and dynamics..." – 7th paragraph)

10. De acordo com o Texto 2, a inteligência comportamental

A. deve ser desenvolvida pelo cientista de dados para que ele seja capaz de manter o controle emocional.

B. pode ser observada a partir da internet das coisas, das mídias sociais e dos serviços de telefonia móvel.

C. pode ser observada a partir das atividades e da dinâmica de indivíduos ou grupos no mundo físico.

D. atua nos sistemas de dados através de intuições, crenças e expectativas.