

Análise de tendências de padrões de escolha usando plataformas de Big Data

Trend Analysis of choice patterns using Big Data platforms

António José Graça, Instituto Politécnico do Cávado e do Ave, Portugal, antonio.costa.graca@gmail.com

Joaquim P. Silva, Instituto Politécnico do Cávado e do Ave, Portugal, jpsilva@ipca.pt

Nuno Lopes, Instituto Politécnico do Cávado e do Ave, Portugal, nlopes@ipca.pt

Resumo

Para sobreviverem atualmente, as organizações necessitam de ser ágeis e de tomar decisões rápidas. As instituições de ensino superior também enfrentam uma elevada competição, o que as obriga a uma constante adaptação aos requisitos do mercado. Estas instituições, todos os anos, são confrontadas com inúmeras questões na definição da estratégia de captação de novos alunos, dos quais estão fortemente dependentes. Com este projeto, pretende-se desenvolver uma solução de suporte à análise dos padrões e tendências de escolha dos alunos colocados no ensino superior, recorrendo às tecnologias de Big Data e mineração de dados.

Palavras-chave: Ensino superior; big data; business intelligence; mineração de dados

Abstract

In the current context, organizations need to be agile and to take rapid decisions. Higher education institutions also face the same competition, so they need to constantly adapt themselves to the market requirements. These institutions, every year, face up numerous questions about the best strategy they must follow for capturing new students. This projects aims to develop a solution for supporting patterns and trends analysis, based in big data and data mining technologies, about the choices the students do in the first time enrolment to higher education.

Keywords: Higher Education; big data; business intelligence; data mining

1. INTRODUÇÃO

Atualmente, num mundo de negócios cada vez mais competitivo, as organizações estão sob constantes pressões e, como consequência, a capacidade de rápida resposta à mudança torna-se um fator crítico de sucesso. Neste contexto, as organizações necessitam de ser ágeis e tomar as decisões de uma forma célere. As instituições do ensino superior (IES), enquanto organizações, ainda que com especificidades próprias do seu ramo de atividade, enfrentam, também elas, as mesmas pressões. No final do século passado, assistiu-se ao aumento significativo de estabelecimentos de ensino superior e, mais recentemente, tem-se verificado uma diminuição do número de candidatos.

Assim, para sobreviverem neste mercado, as IES necessitam de ter um conhecimento cada vez mais profundo dos candidatos e das motivações que condicionam as suas escolhas [Tavares 2011]. Para perceber esta motivação não basta analisar os resultados do concurso nacional de acesso. Neste projeto, pretende-se também analisar a informação sobre as IES e as suas ofertas formativas que é disponibilizada, no contexto

dos sistemas de apoio à decisão e recorrendo a ferramentas usadas atualmente na implementação desses sistemas.

Há cerca de duas décadas, a adoção de sistemas de apoio à decisão pelas organizações registou um acentuado crescimento. Nessa data, contribuiu para esse efeito a existência e o crescimento de repositórios de dados nas organizações, a redução dos custos de computação e o aumento da competição no mercado que se verificou na altura [Berry e Linoff 1997]. Para Inmon [2005], foram fundamentais o surgimento de dispositivos de armazenamento de acesso direto e o amadurecimento dos sistemas de gestão de bases de dados relacionais.

Atualmente, existem no mercado muitas tecnologias para implementar sistemas de apoio à decisão, comercialmente apresentadas como *business intelligence* ou *business analytics*. O principal objetivo dos sistemas de *business intelligence* é disponibilizar ferramentas e metodologias que permitam às organizações tomar decisões efetivas e em tempo útil [Vercellis 2009]. Contudo, os sistemas atuais não lidam muito bem com dados não estruturados, tendo surgido recentemente muitas ferramentas para desenvolver soluções de suporte à decisão que lidam com *big data*, i.e., dados heterogêneos e massivos. Wu et al [2014] propuseram um enquadramento de mineração de dados sobre *big data* com três camadas concêntricas: a camada externa consiste na computação e acesso aos dados; a camada intermédia na privacidade e domínio de aplicação e conhecimento da *big data*; e a camada interna nos algoritmos de mineração de dados sobre *big data*.

Na secção seguinte do artigo, procura-se dar uma visão do ensino superior em Portugal. De seguida, é realizado um breve estudo do estado da arte sobre a análise de tendências de escolha. Na quarta secção, apresenta-se uma abordagem sucinta da arquitetura, fontes de dados e tecnologias a serem usadas neste trabalho. Por último, são identificados os desafios a vencer e os contributos que daí resultarão.

2. ENSINO SUPERIOR EM PORTUGAL

Ao longo das últimas décadas, temos assistido a enormes mudanças no sistema de ensino superior em Portugal. A partir de 1974, registou-se uma grande expansão e diversificação para a qual contribuíram o aumento do número das universidades públicas, a criação de um subsistema politécnico público e a abertura à iniciativa privada. O sistema cresceu e diversificou-se com o aparecimento de novas instituições, novos programas e áreas de ensino e formação, e a entrada de estudantes de origens socioeconómicas diversificadas, que apresentam diferentes situações perante o mercado de trabalho.

O Quadro 1 mostra o número de estabelecimentos de ensino superior por região e estabelecimento de ensino, referente aos anos de 1990 e de 2015. Estes dados são demonstrativos do crescimento acentuado da oferta de ensino superior que se verificou nos últimos anos.

TERRITÓRIOS ANOS	SUBSISTEMA DE ENSINO					
	PÚBLICO		PRIVADO		TOTAL	
	1990	2015	1990	2015	1990	2015
Portugal	125	176	55	117	180	293

Continente	120	169	54	115	174	284
Norte	32	44	22	56	54	100
Centro	31	44	6	13	37	57
Área Metropolitana de Lisboa	41	54	25	41	66	95
Alentejo	11	18	1	2	12	20
Algarve	5	9	0	3	5	12
Região Autónoma dos Açores	3	5	0	0	3	5
Região Autónoma da Madeira	2	2	1	2	3	4

Quadro 1 – Número de estabelecimentos de ensino superior (Fonte: DGEEC/MEC, PORDATA)

A Figura 1 mostra a evolução do número de alunos matriculados pela primeira vez no ensino superior por ano e subsistema de ensino, segundo dados publicados pela PORDATA¹. O valor máximo foi atingido em 2003 quando ultrapassou ligeiramente o valor de 400.000 alunos matriculados.

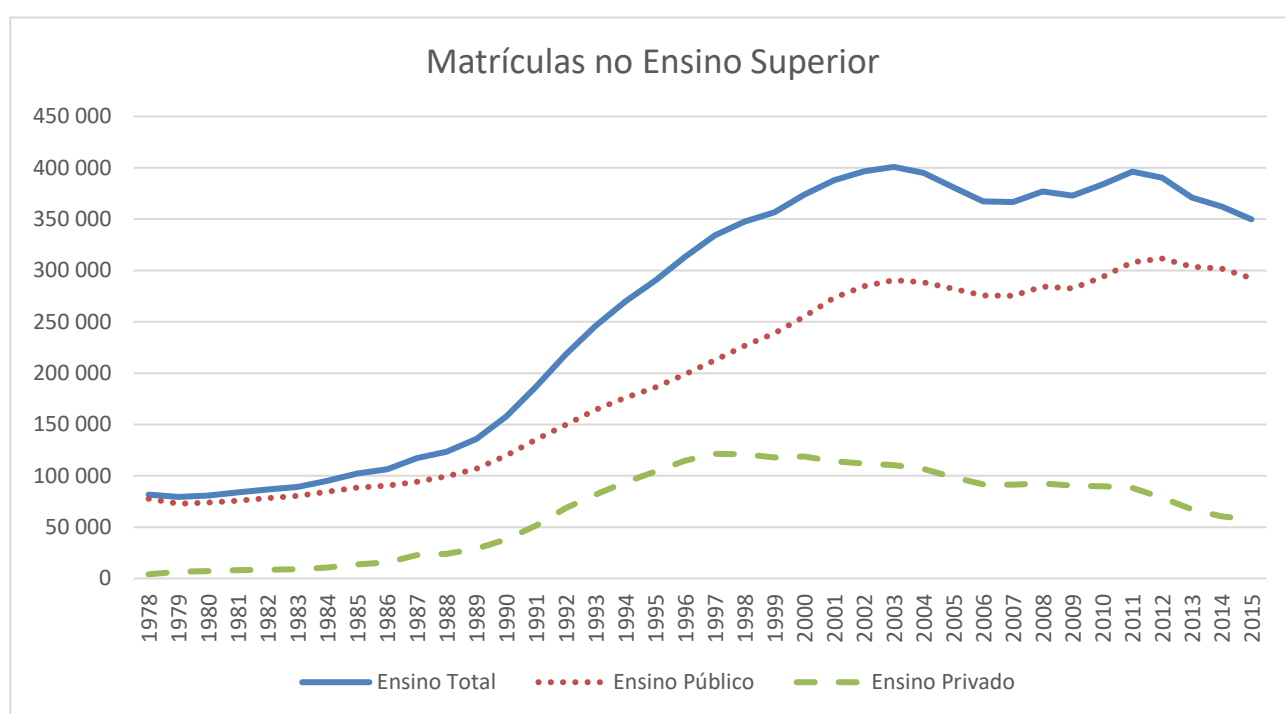


Figura 1 – Alunos matriculados no ensino superior por subsistema de ensino (Fonte: DGEEC/MEC, PORDATA)

Podemos constatar que houve um aumento significativo do número de estabelecimentos de ensino superior entre 1990 e 2015, no valor de 62,78%, e que o número total de alunos matriculados pela primeira vez apresentam uma tendência decrescente, tendo diminuído 15% entre 2011 e 2014.

Entre os vários fatores que contribuem para esta diminuição, destaca-se a diminuição da taxa de natalidade no nosso país, tendência que se verifica desde o início da década 60 e que, segundo dados da PORDATA, atingiu o seu valor mínimo em 2014, com o nascimento de menos de 8 bebés por cada 1000 habitantes.

¹ <http://www.pordata.pt/>

O ensino superior em Portugal é seletivo, pois é regulado por *numerus clausus*. Portanto, as escolhas dos estudantes não são totalmente livres, pois dependem, por um lado, do número de vagas e, por outro, do seu sucesso académico. Os estudantes escolhem os cursos de ensino superior em função da perceção que têm do mercado de trabalho e do ensino superior, o que resulta de fatores socioculturais [Tavares 2015]. Além disso, é possível verificar a existência de uma dimensão espacial implícita nas escolhas que os estudantes fazem quando acedem ao ensino superior e que está muito relacionada com a distribuição espacial das instituições que oferecem cursos de ensino superior [Fonseca e Encarnação 2012] [Sá 2015].

Perante os constrangimentos já referidos e pela análise dos estudos efetuados, torna-se legítimo que se coloquem às IES algumas questões: quais os cursos escolhidos pelos alunos colocados? Quais os de primeira opção? Existe alguma tendência regional? Qual o número de alunos que não foram colocados? O aluno escolhe sempre o mesmo tipo de curso nas diferentes instituições para onde concorre? Existe alguma correlação entre o curso que o aluno escolhe e outros cursos que coloca como opção? Ao longo dos anos, o valor da distância média entre a residência do aluno e o estabelecimento de ensino aumentou ou diminuiu?

Para dar resposta a estas e outras perguntas, surge a necessidade efetuar a consulta e a análise de dados, de diferentes bases de dados, das redes sociais e das notícias referentes aos cursos. Dessa forma, recorrendo a ferramentas de suporte à decisão sobre *big data*, pretende-se determinar tendências de padrões de escolha dos alunos colocados no ensino superior.

3. ANÁLISE DE TENDÊNCIAS DE ESCOLHA

Num estudo realizado pela Noel-Levitz, Inc. em 2011, EUA, onde se procura determinar os fatores que influenciam a escolha dos estudantes que pela primeira vez concorrem ao ensino superior, verifica-se que os principais fatores que influenciam a decisão são o custo, a existência de bolsa de estudos e a reputação da instituição.

De acordo com Ming [2014], devido à forte concorrência no sector do ensino superior na Malásia muitas das universidades tiveram de se tornar mais "empresariais" e implementar estratégias de marketing no recrutamento de estudantes, tanto localmente como no exterior. Nesse estudo, foi proposto desenvolver uma estrutura conceitual para explorar os fatores que influenciam a decisão de escolha da faculdade dos alunos na Malásia. As variáveis independentes que foram identificadas como tendo influenciado a decisão da escolha da faculdade dos alunos foram a localização, o programa académico, a reputação de faculdade, as instalações, o custo, a disponibilidade de ajuda financeira, as oportunidades de emprego, a publicidade, as próprias Instituições de Ensino Superior (IES), os representantes e a visita ao campus.

No caso do ensino superior Português, a escolha da instituição de ensino pode também estar relacionada com a decisão de sair de casa, existindo diferenças entre géneros no tipo de escolha da instituição de ensino superior. O background socioeconómico também parece restringir as escolhas dos estudantes e a acessibilidade desempenha um papel nas suas decisões [Sá et al 2011].

4. ABORDAGEM

As escolhas dos estudantes, como vimos anteriormente, estão condicionadas por múltiplos fatores. O tipo das escolhas e a necessidade de cada vez mais as IES terem uma gestão “empresarial”, implicam a consulta de diferentes bases de dados como as da DGES, do INE, do IIEFP, a obtenção de informação das redes sociais, assim como da informação disponibilizada nos sites com notícias sobre os cursos e as IES.

Pretende-se recolher os todos os dados para uma plataforma de *big data*, os quais deverão ser processados realizando um ciclo simplificado do processo de mineração de dados, segundo a metodologia CRISP-DM, conforme se apresenta na Figura 2: (a) identificação do tipo de padrões a procurar; (b) preparação dos dados (pré-processamento) recorrendo a ferramentas de *big data*; (c) desenvolvimento do modelo recorrendo a ferramentas de *data mining*; e (d) análise e validação dos resultados.

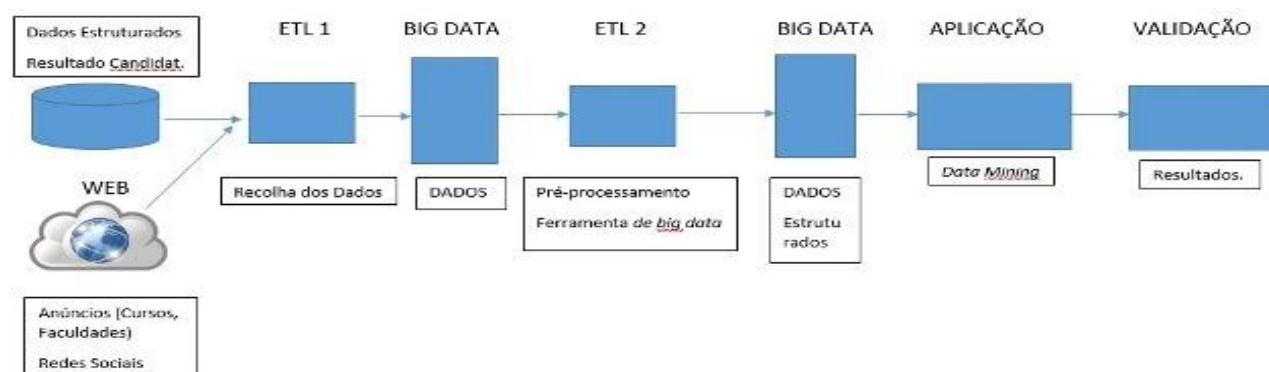


Figura 2 – Representação simplificado do processo a adotar no estudo

Para as organizações obterem valor dos seus dados têm de se apoiar em tecnologias e sistemas de análise de dados. Cada vez mais, para a tomada de decisão são integrados dados heterogêneos provenientes de fontes externas. *Big data* é um termo amplo para conjuntos de dados muito grandes ou complexos que tornam inadequadas as aplicações de processamento de dados tradicionais. Os desafios incluem todo o processo de desde a obtenção dos dados: análise, captura, tratamento de dados, pesquisa, partilha, armazenamento, transferência, visualização e consulta.

5. CONCLUSÕES E TRABALHO FUTURO

Como se constatou existem vários fatores que condicionam e determinam a escolha dos candidatos. Os estudos analisados até agora normalmente incidem sobre algum desses fatores não contemplando, simultaneamente, todos eles. Este estudo irá preparar uma plataforma de *big data* para a recolha de dados das diferentes fontes de dados que permitam elaborar uma análise conjunta de todos os condicionantes, políticos, económicos, demográficos, culturais e sociais, de forma a permitir o conhecimento do padrão de escolha dos candidatos.

Através de uma análise de padrões com a informação recolhida, serão identificadas tendências e os fatores que as influenciam.

6. REFERÊNCIAS

- Berry, M.J.A. and Linoff, G.S. (1997). *Data Mining Techniques: For Marketing, Sales and Customer Support*. John Wiley & Sons.
- Inmon, W. H. (2005). *Building the Data Warehouse*, 3rd Edition, John Wiley & Sons.
- Fonseca, M., & Encarnação, S., “A Massificação do Ensino Superior em Portugal: Efeitos Espaciais na Diversificação do Sistema”, XIII Coloquio Ibérico de Geografia, Santiago De Compostela, (outubro de 2012)
- Ming, J. S. K., *Institutional Factors Influencing Students' College Choice Decision in Malaysia: A Conceptual Framework*, (2010)
- Sá, Carla, et al. "Higher education (related) choices in Portugal: joint decisions on institution type and leaving home." *Studies in Higher Education* 36.6 (2011): 689-703.
- Sá, Carla. "Mobilidade Geográfica e Acesso ao Ensino Superior: Padrões e Determinantes." *Acesso ao Ensino Superior: Desafios para o Século XXI, Coleção: Seminários e Colóquios*. Edição Eletrónica. ISBN (2015): 978-972.
- Tavares, O. "Acesso ao Ensino Superior: As Preferências e a Consistência das Escolhas do Estudantes." *Acesso ao Ensino Superior: Desafios para o Século XXI, Coleção: Seminários e Colóquios*. Edição Electrónica. ISBN (2015): 978-972.
- Tavares, O. *As Escolhas dos Estudantes no Acesso ao Ensino Superior Português—processos e racionalidades*. Tese de Doutoramento. Porto: Faculdade de Psicologia e Ciências da Educação da Universidade do Porto, 2011.
- Vercellis, C., *Business Intelligence: Data Mining and Optimization for Decision Making*. Wiley-Blackwell, (2009).
- Wu, Xindong, et al. "Data mining with big data." *Knowledge and Data Engineering, IEEE Transactions on* 26.1 (2014): 97-107.