

O PROCESSO *KDD* NO ESTUDO DO PERFIL DOS INGRESSANTES/2012 NA UNIVERSIDADE ESTADUAL DO OESTE DO PARANÁ – *CAMPUS CASCAVEL*

Mychelly Agnes Marcelo Henrique¹
Rosângela Villwock²

Área de conhecimento: Administração/ Matemática
Eixo Temático: Administração de Pessoas, comportamento, Gestão do
Conhecimento/ Mineração de Dados

RESUMO

Este trabalho consiste na aplicação do processo *KDD* – *Knowledge Discovery in Database*, que significa “Descoberta de conhecimento em base de dados”, visando conhecer o perfil socioeconômico dos acadêmicos ingressantes/2012 na Universidade Estadual do Oeste do Paraná (UNIOESTE) – *Campus Cascavel*. Os dados utilizados foram obtidos de questionários respondidos pelos acadêmicos no momento da inscrição para o vestibular. A Análise de Agrupamentos gerou grupos de acadêmicos nivelados de acordo com as suas características socioeconômicas, possibilitando conhecimentos que poderão ser utilizados para subsidiar decisões dentro da Universidade. A Análise de Agrupamentos foi realizada com o auxílio do *software Statgraphics Centurion XVI.1* versão demonstrativa e como resultado obteve-se cinco grupos que não se caracterizaram por um centro apenas. Analisando-se os resultados verificou-se que o perfil dos acadêmicos da UNIOESTE – *Campus Cascavel* – não revelam uma relação direta entre as condições socioeconômicas e a escolha do curso, ao contrário, observou-se acadêmicos de um mesmo curso pertencendo a diferentes grupos socioeconômicos. Além disso, foi possível constatar que os cursos mais procurados alcançam acadêmicos dos diversos grupos.

Palavras-chave: Agrupamento. Mineração de Dados. Perfil Socioeconômico.

INTRODUÇÃO

Segundo Uemura (2010) a Universidade é uma organização complexa e administrá-la é uma tarefa difícil, que poucos fazem com competência e desembaraço. A sua eficácia pode ser verificada no desempenho da Instituição e sua eficiência, nos resultados. Os conhecimentos teóricos construídos nos bancos escolares não são suficientes para administrá-la, pois para isso é necessário colocar em prática o exercício dos conhecimentos e das experiências e ter clareza daquilo que se quer, além de requerer um conhecimento dos meios dos quais dispõe e/ou necessita.

Ainda de acordo com o autor é importante conhecer a universidade e o seu contexto socioeconômico, uma vez que só assim será possível identificar as

¹ Graduanda em Matemática – mychelly_agnes@hotmail.com.

² Doutora em Métodos Numéricos em Engenharia – rosangela.villwock@unioeste.br.



mudanças necessárias e proporcionar instrumentos para tomadas de decisões mais acertadas em busca de uma universidade de qualidade ou, nas palavras do próprio autor “conhecer a universidade é preciso, mas conhecer o seu contexto sócio-econômico-cultural é indispensável” (UEMURA, 2010, p. 91).

De acordo com Carvalho (2006), os dados sobre o perfil dos estudantes no ensino superior revelam que a democratização do ensino é bastante complexa no Brasil, dada a brutal desigualdade de renda entre as famílias e a reduzida oferta de ensino gratuito e de qualidade. Segundo Lage (2007), no Brasil, dos jovens que concluem o Ensino Médio, apenas 11% conseguem chegar à universidade, sendo que a maioria deles proveniente da rede privada de ensino.

Segundo Martinhago (2005), em busca de melhores níveis de ensino, a maioria das instituições solicitam o preenchimento de uma ficha com um questionário socioeconômico e cultural como forma de obter informações que possam levar ao conhecimento sobre os candidatos ao processo seletivo para o ensino superior. Esses dados podem auxiliar os administradores das instituições na tomada de decisões, a fim de melhorar a qualidade de ensino.

Martinhago (2005), que realizou um estudo sobre o processo seletivo do vestibular da UFPR, visando apresentar a relação existente entre as variáveis socioeconômicas com o desempenho dos candidatos, sugere que os resultados obtidos devem ser utilizados pela universidade para definir novas regras para os próximos vestibulares e assim, implementar ações que visem melhorar a qualidade de ensino, diminuindo a evasão e direcionando o candidato ao vestibular na escolha do curso baseando-se no seu perfil. O autor fez um estudo através de mineração de dados aplicados ao processo seletivo do vestibular da UFPR e verificou que, nos cursos mais concorridos, os dados socioeconômicos e culturais do candidato são relevantes para o seu desempenho e que o mesmo não ocorre nos cursos menos concorridos. As pontuações obtidas nas provas de outras áreas, diferentes daquelas a qual o curso pertence, influenciaram na aprovação dos candidatos a cursos das áreas de exatas e humanísticas. Observou-se ainda que a nota de língua estrangeira é um atributo importante no sucesso do candidato em praticamente todos os cursos.

Segundo Sousa, Oliveira e Nascimento (2009), que analisaram o desempenho acadêmico dos alunos da Universidade Federal do Piauí (UFPI)



campus Ministro Petrônio Portela, por meio de análise socioeconômica, constatou-se um perfil diferente para estudantes de centros diferentes. Para os autores, conhecer o perfil socioeconômico dos alunos é muito importante para a universidade, pois permite ter uma ideia de quais as condições de vida dos estudantes, bem como desenvolver políticas públicas que se encaixem com o perfil encontrado. Os autores revelaram que os estudantes do *campus* apresentam uma quantidade de reprovações relativamente baixa para a quantidade de horas de estudo informada. Através de uma análise por sexo observaram que as mulheres possuem menos reprovações, são mais jovens (em média) e estudam quase uma hora a menos que os homens.

Oliveira (2011) realizou um estudo na Universidade Federal da Paraíba (UFPB), sobre como as características socioeconômicas dos estudantes podem determinar seu desempenho acadêmico. Utilizando o questionário socioeconômico observou que é fundamental para o estudante possuir uma base familiar para que possa garantir seu melhor desempenho em uma instituição de Ensino Superior. Entretanto, ainda segundo a autora, a Instituição também poderia adotar políticas que controlassem a evasão, como por exemplo, mudanças no regimento interno sobre trancamentos e reprovações e um acompanhamento mais efetivo por parte da coordenação dos cursos, de modo a estimular os alunos ao término do curso em tempo hábil.

Moreira *et al.* (2012) realizou um estudo sobre o perfil dos ingressantes nos cursos de licenciatura em matemática no Brasil, com dados obtidos através de um questionário sobre condições socioeconômicas, formação escolar e processo de escolha da licenciatura em matemática como formação universitária, respondido por 664 alunos que cursam licenciatura em matemática, em 19 instituições de ensino superior, de 10 estados brasileiros, nos anos de 2008 a 2010. Os resultados indicam que, em sua grande maioria, o ingressante é jovem (abaixo de 25 anos), solteiro, estudou na escola pública, escolheu a licenciatura atraído mais pela matemática do que pela docência, possui renda familiar abaixo de 5 salários mínimos, não contribui para o sustento da família e está ascendendo a um nível de escolaridade superior ao dos pais. Além disso, não há concentração do sexo feminino, como no caso dos docentes em geral e 55% dos estudantes não estão seguros se irão exercer efetivamente a profissão, ao se graduarem.



Spíndola, Martins e Francisco (2008) compararam o perfil socioeconômico com fatores que interferem na opção de escolha pela enfermagem, de alunos de duas instituições de ensino no município do Rio de Janeiro. Concluíram que a maioria dos que se interessam pelo curso são adultos jovens, mulheres, solteiras e pertencentes às classes B e C. O mercado de trabalho em expansão e a oportunidade de uma melhor remuneração são fatores que atraem os alunos. Além disso, segundo os autores a presença de familiares que atuem na área da saúde é um dos fatores que contribuem para a escolha. Ações voltadas para a orientação de alunos de Ensino Médio poderiam esclarecê-los sobre a profissão e o mercado de trabalho, despertando assim o seu interesse.

Diante do exposto, este trabalho pretende encontrar respostas a questionamentos como: Existem características ou fatores socioeconômicos que influenciam o aluno na escolha do curso?; A universidade, de posse destas informações, pode desenvolver estratégias para que estes fatores não sejam determinantes para o ingresso e permanência do acadêmico?; Existem projetos que possam ser desenvolvidos associados ao perfil dos acadêmicos?; Para carreiras de grande prestígio social ou que projetam a imagem de possibilitar elevados rendimentos, verifica-se um maior número de acadêmicos com situação socioeconômica mais favorecida?; Baseando-se no perfil dos acadêmicos, é possível desenvolver estratégias para diminuir a evasão?.

Deste modo, teve-se a iniciativa de pesquisar os dados sócio-econômicos dos ingressantes na UNIOESTE – *Campus Cascavel* em 2012, com o intuito de conhecer as características relevantes dos ingressantes em busca de indicadores para formular políticas, vislumbrando atendê-los conforme suas necessidades. Este estudo visa oferecer subsídios para a elaboração de projetos que contribuam para a permanência dos acadêmicos, melhorando a realidade da Universidade no que diz respeito a evasão, por exemplo. Estudar o perfil sócio-econômico permite identificar as mudanças necessárias, proporcionar instrumentos para complementar a tomada de decisões pelos profissionais responsáveis e definir ações para melhorar a Universidade.

O objetivo geral do presente trabalho foi delinear o perfil dos acadêmicos ingressantes/2012 na Universidade Estadual do Oeste do Paraná (UNIOESTE) – *Campus Cascavel*, através do processo de Descoberta de Conhecimento em Bases



de Dados, utilizando a Análise de Agrupamentos.

O texto que segue está organizado da seguinte forma: na seção Referencial Teórico são apresentadas as descrições do Processo *KDD*, das tarefas de Mineração de Dados e da técnica que será utilizada neste trabalho; na seção Metodologia são apresentados os dados utilizados no trabalho e o algoritmo utilizado na tarefa de Agrupamento; na seção seguinte são apresentados e analisados os resultados, seguidos das Considerações Finais.

1 REFERENCIAL TEÓRICO

O mundo tem armazenado uma quantidade incrível de dados que superam em muito as nossas habilidades para interpretá-los e digeri-los, criando a necessidade de geração de ferramentas e técnicas para automatizar e analisar a base de dados de forma inteligente (FAYYAD *et al.*, 1996).

Essas técnicas e ferramentas que buscam transformar esses dados armazenados em conhecimento são objetos do campo emergente chamado *Knowledge Discovery in Databases – KDD* (que significa “Descoberta de Conhecimento em Base de Dados”) (TAN; STEINBACH; KUMAR, 2009).

Segundo Fayyad *et al.* (1996), trata-se de um processo não trivial de descoberta de padrões válidos, novos, úteis e acessíveis. A principal vantagem do processo de descoberta é que não são necessárias hipóteses, sendo que o conhecimento é extraído dos dados sem conhecimento prévio. O processo *KDD* é o conjunto de atividades contínuas que compartilham o conhecimento descoberto a partir de bases de dados, um dos objetivos do *KDD* é tomar os padrões assimiláveis ao conhecimento humano. O processo *KDD* consiste em cinco etapas, ilustradas na Figura 1.

Na **Seleção**, os dados a serem utilizados são selecionados de acordo com os critérios definidos (FAYYAD *et al.*, 1996).

No **Pré-processamento**, os dados de entrada são modificados para um formato apropriado para as análises subsequentes. As informações julgadas desnecessárias são removidas e há a reconfiguração dos dados para assegurar formatos consistentes (FAYYAD *et al.*, 1996).

Segundo Tan, Steinbach e Kumar (2009), os passos do Pré-processamento incluem: Limpeza dos dados; Observações duplicadas; Seleção de registros e



características que sejam relevantes à tarefa de mineração de dados.

Na limpeza dos dados e observações duplicadas, retiram-se os dados que não são considerados relevantes para a pesquisa e então é feita uma limpeza nas questões duplicadas, por isso segundo o autor o pré-processamento é o passo mais trabalhoso e demorado no processo geral de descoberta de conhecimento.

Na **Transformação ou Formatação**, os dados são transformados em formatos utilizáveis, usando métodos de redução de dimensionalidade dos dados, conforme a técnica de Mineração de Dados, (TAN; STEINBACH; KUMAR, 2009).

De acordo com Tan, Steinbach e Kumar (2009) a utilização de métodos de redução de dimensionalidade são utilizados, porque muitos dos algoritmos de mineração de dados funcionam melhor se o número de atributos nos dados for menor. Além disso, muitos algoritmos tem problemas com dados de alta dimensionalidade.

Na **Mineração de dados** (MD), esta etapa é o núcleo do processo, onde são aplicados os algoritmos para extrair padrões dos dados (FAYYAD *et al.*, 1996).

Na **Avaliação**, depois de identificados os padrões pelo sistema, estes são interpretados, gerando conhecimentos através de informações antes escondidas nos dados, que darão suporte à tomada de decisões (FAYYAD *et al.*, 1996). O objetivo de interpretar os resultados é filtrar as informações que serão apresentadas a aqueles responsáveis em tomar decisões. (TAN; STEINBACH; KUMAR, 2009).

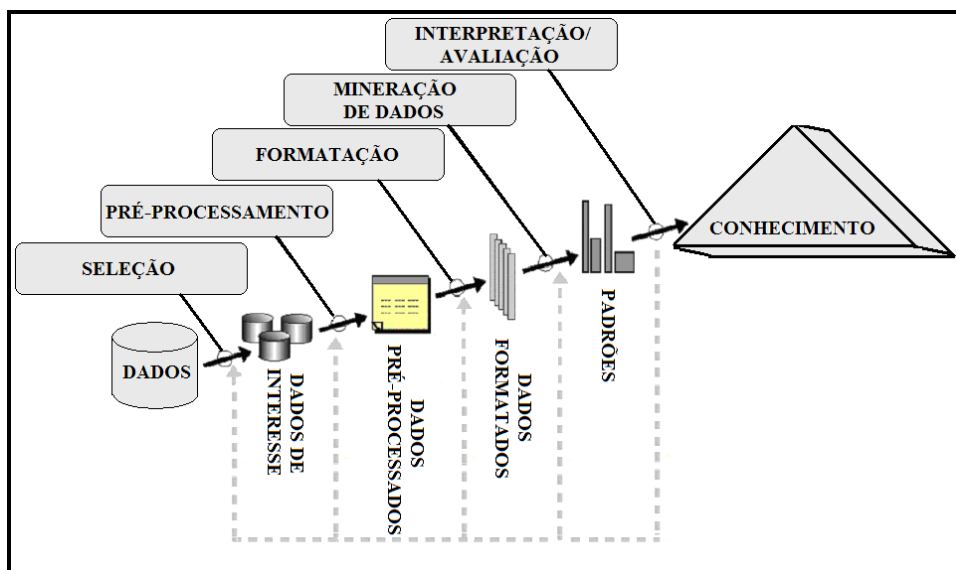


Figura 1 – Etapas do processo KDD, adaptada de Fayyad *et al.* (1996).

O campo da Mineração de Dados saiu dos limites das técnicas atuais de



análise de dados. É uma tecnologia que combina métodos tradicionais de análise de dados com algoritmos sofisticados para processar grandes volumes de dados, que podem ser usadas para apoiar uma ampla gama de aplicações de inteligência de negócio, como a criação de perfil de clientes, vendas direcionadas, detecção de fraudes e também nas áreas de medicina, ciências e engenharia. As técnicas de mineração de dados são organizadas para agir sobre grandes bancos de dados com o intuito de descobrir padrões úteis que poderiam de outra forma, permanecer ignorados (TAN; STEINBACH; KUMAR, 2009).

As principais tarefas de Mineração de Dados estão relacionadas à Classificação, Associação e Agrupamento de padrões (FAYYAD *et al.*, 1996).

Na **Classificação** cada padrão contém um conjunto de atributos e um dos atributos é denominado classe. O objetivo da classificação é encontrar um modelo para predição da classe como função dos outros atributos. É a tarefa de organizar objetos em uma entre as diversas categorias pré-definidas e é um problema universal que engloba muitas aplicações diferentes. As técnicas de classificação são mais apropriadas para prever ou descrever um conjunto de dados com categorias nominais ou binárias, elas são menos efetivas para categorias ordinais (TAN; STEINBACH; KUMAR, 2009).

Na **Associação** o objetivo é produzir regras de dependência que irão prever a ocorrência de um atributo baseado na ocorrência de outros atributos (FAYYAD *et al.*, 1996).

Segundo, Tan, Steinbach e Kumar (2009) a associação é útil para descobrir relacionamentos interessantes escondidos em um conjunto grande de dados. Muitos varejistas se interessam em analisar os dados para aprender sobre o comportamento de compra de seus clientes.

No **Agrupamento** ou Segmentação (*Clustering*) procuram-se grupos de padrões tal que padrões pertencentes a um mesmo grupo são mais similares uns aos outros e dissimilares a padrões em outros grupos. (TAN; STEINBACH; KUMAR, 2009).

De acordo com Hair Jr *et al.* (2005), a análise de agrupamentos é o nome que se dá a um grupo de técnicas utilizadas para desenvolver subgrupos significativos de objetos e seu objetivo é classificar os objetos em um pequeno número de grupos mutuamente excludentes.



Agrupamento de Dados

Os objetos são agrupados segundo informações encontradas nos dados que o descrevem, bem como seus relacionamentos. O objetivo é que os objetos dentro de um grupo sejam semelhantes, portanto quanto maior a semelhança dentro de um grupo, e maior a diferença entre os grupos, melhor ou mais distinto será o agrupamento. Na análise de grupos, os grupos de dados se dividem de forma que tenham significados e em alguns casos a análise de grupos é apenas um ponto inicial útil para outros propósitos, como resumo de dados. Deste modo, seja para a compreensão ou utilidade, a análise de grupos tem há muito desempenhado um papel importante em uma ampla variedade de campos: psicologia, biologia, estatística, reconhecimento de padrões, recuperação de informações e mineração de dados, conforme afirmam (TAN; STEINBACH; KUMAR, 2009).

Segundo Tan, Steinbach e Kumar (2009), os algoritmos de agrupamento podem ser hierárquicos ou de particionamento (não hierárquico).

Os métodos hierárquicos englobam técnicas que buscam de forma hierárquica os grupos e, por isso, admitem obter vários níveis de agrupamento. Os métodos hierárquicos podem ser subdivididos em aglomerativos e divisivos (FAYYAD *et al.*, 1996).

O método hierárquico aglomerativo considera, a princípio, cada padrão como um grupo e agrupa o par de grupos com maior similaridade em um novo grupo até formar um único grupo contendo todos os padrões (TAN; STEINBACH; KUMAR, 2009).

O método hierárquico divisivo, ao contrário, inicia com um único grupo e executa um processo de sucessivas subdivisões (DINIZ; LOUZADA-NETO, 2000).

Os métodos de agrupamentos hierárquicos mais conhecidos são: Ligação simples, ligação completa, ligação média e método de ward (FAYYAD *et al.*, 1996).

Os métodos não-hierárquicos ou de particionamento procuram uma partição sem a necessidade de associações hierárquicas. Seleciona-se uma partição dos elementos em k grupos, otimizando algum critério (DINIZ; LOUZADA-NETO, 2000). É uma divisão do conjunto em grupos não interseccionados de modo que cada objeto de dado esteja exatamente em um subconjunto (TAN; STEINBACH; KUMAR, 2009).



Os grupos podem ser baseados em protótipos, para dados com atributos contínuos, o protótipo de um grupo é muitas vezes um centróide, que geralmente é a média de todos os pontos de um grupo (TAN; STEINBACH; KUMAR, 2009).

O método de agrupamento não-hierárquicos mais conhecido é o *K-means* (*K*-médias) (FAYYAD *et al.*, 1996). Este método será utilizado neste trabalho.

2 METODOLOGIA

Segundo UNIOESTE (2012), o número total de cursos contabiliza 33 graduações, 15 delas no *Campus* de Cascavel. O *Campus* de Cascavel conta com 3.219 alunos cursando a graduação em 2012, divididos em 21 turmas.

A Universidade Estadual do Oeste do Paraná *campus* Cascavel possuía setecentos e dez (710) vagas para o *Campus* de Cascavel para ingresso em 2013 distribuídas entre os dezoito (18) cursos de graduação, divididos em cinco (05) centros:

- **CCBS** (Centro de Ciências Biológicas e da Saúde), ao qual pertencem os cursos de Ciências Biológicas, Enfermagem, Fisioterapia e Odontologia;
- **CCMF** (Centro de Ciências Médicas e Farmacêuticas), ao qual pertencem os cursos de Farmácia e Medicina;
- **CCET** (Centro de Ciências Exatas e Tecnologias), ao qual pertencem os cursos de Ciência da Computação, Engenharia Agrícola, Engenharia Civil e Matemática;
- **CCSA** (Centro de Ciências Sociais Aplicadas), ao qual pertencem os cursos de Administração, Ciências Contábeis e Ciências Econômicas;
- **CECA** (Centro de Educação, Comunicação e Artes), ao qual pertencem os cursos de Letras – Português/Inglês, Letras – Português/Espanhol, Letras – Português/Italiano, Pedagogia Matutino e Pedagogia Noturno.

A pesquisa foi feita com base em dados disponibilizados pela universidade coletados via questionários sócio-econômicos, aplicados aos vestibulandos 2011/2012. Neste sentido foram analisados os dados relativos aos acadêmicos que ingressaram na universidade em 2012, no *campus* de Cascavel, nos 18 cursos ofertados, totalizando 711 acadêmicos.

Na confecção da base de dados algumas questões foram adaptadas



(subdivididas ou excluídas) totalizando 30 questões analisadas. Algumas questões não foram utilizadas, pois não foram consideradas relevantes para o objetivo deste trabalho. Além disso, a questão sobre qual o tipo de necessidade especial possui não foi utilizada, pois seria possível identificar o portador de tal necessidade.

Neste trabalho a tarefa de MD é o Agrupamento. A Análise de Agrupamentos foi realizada com o auxílio do *software Statgraphics Centurion XVI.1* versão demonstrativa. Foi aplicado o método *K*-médias, cujo algoritmo pode ser consultado em Tan, Steinbach e Kumar (2009). A medida de dissimilaridade adotada foi a Distância Euclidiana.

Segundo Tan, Steinbach e Kumar (2009), *K*-médias é uma técnica particional de agrupamento baseadas em protótipos que tenta encontrar *m* número especificado pelo usuário de grupos “*k*”, representados pelos seus centróides, que geralmente é a média de um grupo de pontos, sendo esse um dos algoritmos de agrupamento mais antigos e amplamente usados.

A técnica é simples, primeiro escolhe-se *K* centróides iniciais (o número de grupos desejados), cada ponto é atribuído a seguir ao centróide mais próximo, e cada coleção de pontos atribuída a um centróide é um grupo, o centróide de cada grupo é então atualizado baseado nos pontos atribuídos ao grupo, então o passo se repete até que nenhum ponto mude de grupo, ou equivalentemente até que os centróides permaneçam os mesmos (TAN; STEINBACH; KUMAR, 2009).

3 APRESENTAÇÃO E ANÁLISE DOS RESULTADOS

Abaixo são apresentados os resultados obtidos com a aplicação da técnica de mineração de dados, mais especificamente a Análise de Agrupamentos, com aplicação do Método *K*-médias.

O *K*-médias foi aplicado considerando-se 5 grupos. Esta quantidade de grupos foi assim escolhida devido à quantidade de centros na Universidade, na tentativa de observar um perfil para cada centro.

Na tabela 1 são apresentados os centroides de cada grupo para as variáveis: Instrução do Pai, Instrução da Mãe, Renda familiar, Casa própria e Naturalidade (Paraná).



Tabela 1 Centróides dos grupos (conjunto de variáveis 1)

Grupos	Instrução do Pai	Instrução da Mãe	Renda Familiar	Casa Própria	Naturalidade (Paraná)
Grupo 1	3,21839	3,6092	2,06897	0,436782	0,908046
Grupo 2	2,94239	3,39095	1,64609	0,37037	0,934156
Grupo 3	3,60714	3,70536	2,20536	0,241071	0,946429
Grupo 4	2,2	2,55455	1,79091	0,518182	0,9
Grupo 5	5,16352	5,54088	3,08805	0,289308	0,811321

Em relação à Instrução do pai é possível observar que o grupo 5 foi o que obteve maior média e o grupo 4 a menor, o mesmo aconteceu em relação a instrução da mãe. No grupo 5, 51,57% dos pais e 63,52% das mães possuem graduação e/ou pós graduação e apenas 10,06% dos pais e 5,66% das mães não concluíram o Ensino Médio. Já no grupo 4, 5,45% dos pais e 10,9% das mães possuem graduação e/ou pós graduação, enquanto 69,09% dos pais e 65,45% das mães não concluíram o Ensino Médio.

Com relação à renda familiar o grupo 5 obteve a maior média e o grupo 2 a menor, seguido pelo grupo 4 com a segunda menor média. O grupo 5 possui 68,55% dos acadêmicos com renda superior a 5 salários mínimos e 6,29% com renda inferior a 2 salários mínimos. Enquanto o grupo 2 possui 14,4% dos acadêmicos com renda superior a 5 salários mínimos e 46,5% com renda inferior a 2 salários mínimos. Já o grupo 4 apresenta 20% dos acadêmicos com renda superior a 5 salários mínimos e 45,45% com renda inferior a 2 salários mínimos.

Em relação a casa própria pode-se observar que o grupo 4 possui a maior média onde 51,81% dos acadêmicos possuem casa própria, seguido do grupo 1 que 43,67% do grupo possui casa própria. O grupo 5 é o que possui a menor média representando 28,93% do grupo com casa própria. No grupo 5, dos 71,07% que não possuem casa própria, 93,02% (ou 65,4% do total do grupo) moram com os pais ou em república. Já no grupo 4, dos 48,19% que não possuem casa própria, 62,26% (ou 30% do total do grupo) moram com os pais ou em república.

Com relação à naturalidade, os grupos apresentaram médias próximas, sendo que o grupo 3 apresenta maior média, onde 94,64% dos acadêmicos são nascidos no Paraná.

Na tabela 2 são apresentados os centróides para cada grupo para as variáveis: Raça Branca, Residentes no Paraná, Trabalha fora, Cursou o Ensino



Fundamental em Escola Pública, cursou o Ensino Médio em Escola Pública.

Tabela 2 Centróides dos grupos (conjunto de variáveis 2)

Grupos	Raça Branca	Residentes no Paraná	Trabalha	Cursou o Ens. Fund. Escola Pública	Cursou o Ens. Médio Escola Pública
Grupo 1	0,816092	1	0,333333	0,91954	0,793103
Grupo 2	0,781893	0,99177	0,296296	0,979424	0,946502
Grupo 3	0,875	0,973214	0,366071	0,973214	0,964286
Grupo 4	0,763636	0,990909	0,881818	0,881818	0,918182
Grupo 5	0,855346	0,924528	0,188679	0,106918	0,0691824

Em relação à Raça os grupos 3 e 5 apresentaram maior média percentual de raça branca em relação aos demais grupos.

Com relação ao local de Residência o grupo 1 apresenta 100% dos acadêmicos residentes no Paraná. O grupo 5 é o que apresenta menor média, sendo que 92,45% dos acadêmicos do grupo são residentes no estado.

Em relação à classe trabalhadora, o grupo 4 é o que possui maior média, e o grupo 5 a menor, seguido pelo grupo 2. No grupo 4, 88,18% dos acadêmicos trabalham, enquanto que no grupo 5 apenas 18,87% dos acadêmicos são trabalhadores e no grupo 2, 29,63% dos acadêmicos trabalham.

Com relação aos Ensinos Fundamental e Médio cursados em Escola Pública, os grupos 2 e 3 apresentam a maior média e o grupo 5 a menor. Enquanto no grupo 2 o Ensino Fundamental foi cursado em Escola Pública por 97,94% e o Ensino Médio por 94,46% dos acadêmicos, no grupo 5 apenas 10,69% dos acadêmicos cursaram o Ensino Fundamental em Escola Pública e 6,91% cursaram o Ensino Médio em Escola Pública.

Na tabela 3 são apresentados os centróides para cada grupo para as variáveis: Cursou o Ensino Médio na Modalidade Regular, Cursou Ensino Médio Educação Geral, Cursou Ensino Médio Magistério, Cursou Ensino Médio Técnico, Cursou Ensino Médio Auxiliar e Cursou Ensino Médio Básico.

Em relação ao Ensino Médio, pode-se observar na tabela acima, que todos os grupos são caracterizados por terem cursado o Ensino Médio na Modalidade regular. Nos grupos 2 e 3, 100% dos acadêmicos cursaram Ensino Médio nesta modalidade. O grupo 4 foi o que teve a menor média, sendo que 82,72% dos acadêmicos do grupo fizeram Ensino Médio Regular.



Tabela 3 Centróides dos grupos (conjunto de variáveis 3)

Grupos	Cursou Ens. Médio na Modalidade Regular	Cursou Ens. Médio Educação Geral	Cursou Ens. Médio Magistério	Cursou Ens. Médio Técnico	Cursou Ens. Médio Auxiliar	Cursou Ens. Médio Básico
Grupo 1	0,988506	0,666667	0,0344828	0,0804598	0	0,218391
Grupo 2	1	0,646091	0,0823045	0,0411523	0	0,230453
Grupo 3	1	0,607143	0,0446429	0,0357143	0	0,321429
Grupo 4	0,827273	0,663636	0,0090909	0,1	0,0181818	0,209091
Grupo 5	0,993711	0,798742	0	0,0125786	0	0,188679

Quanto ao curso feito no Ensino Médio observa-se que todos os grupos são caracterizados por cursar Educação Geral. O grupo 5 é o que apresenta a maior média, com 79,87% dos acadêmicos. Nos demais grupos, em torno de 60% dos acadêmicos também cursaram o Ensino Médio Educação Geral. No grupo 2, 8,23% dos acadêmicos fizeram Magistério. Já o grupo 4, 10% dos acadêmicos cursaram Ensino Médio Técnico e 1,82% dos acadêmicos cursaram Ensino Médio Auxiliar. No grupo 3, 32,14% dos acadêmicos cursaram o Ensino Médio Básico.

Na tabela 4 são apresentados os centróides para cada grupo para as variáveis: Cursou a maior parte do Ensino Médio período diurno, Frequentou cursinho preparatório para o vestibular, Já iniciou algum curso superior.

Tabela 4 Centróides dos grupos (conjunto de variáveis 4)

Grupos	Cursou Ens. Médio Maior parte período diurno	Frequentou cursinho preparatório para o vestibular	Já havia iniciado algum curso superior
Grupo 1	0,896552	0,62069	0,183908
Grupo 2	0,901235	0,444444	0,0493827
Grupo 3	0,973214	0,544643	0,25
Grupo 4	0,518182	0,263636	0,554545
Grupo 5	0,993711	0,515723	0,201258

Com relação ao Período cursado no Ensino Médio, pode-se observar que o grupo 5 possui maior média no período diurno, representando 99,37% do grupo, enquanto o grupo 4 possui apenas 51,81% que cursaram o Ensino Médio neste período.

Em relação aos que frequentaram cursinho preparatório para o vestibular, observa-se que o grupo 1 apresenta a maior média representando 62,07% do grupo. Já o grupo 4 a menor com 26,36% do grupo que frequentaram cursinho.



Analisando os que já haviam iniciado algum curso superior anteriormente o grupo 4 se destaca com 55,45%, enquanto o grupo 2 apenas 4,93% do grupo já havia frequentado algum curso superior.

Nas tabelas 5 e 6 são apresentados os centróides para cada grupo para a variável: Principal motivo de prestar vestibular na UNIOESTE.

Tabela 5 Centróides dos grupos (conjunto de variáveis 5)

Principal motivo de prestar vestibular na UNIOESTE				
Grupos	Atende as condições sócioeconômicas da família	Não posso me deslocar para outro lugar	Ela possui o curso que desejo cursar	O curso escolhido é de reconhecida qualidade
Grupo 1	0	0	0	1
Grupo 2	0,967078	0,00823045	0	0
Grupo 3	0	0,0625	0,803571	0
Grupo 4	0,581818	0,0363636	0,0181818	0,0363636
Grupo 5	0,320755	0,0566038	0,333333	0,150943

Tabela 6 Centróides dos grupos (conjunto de variáveis 6)

Principal motivo de prestar vestibular na UNIOESTE				
Grupos	Oferece curso noturno então posso trabalhar durante o dia	Oferece bom número de cursos, permitindo uma escolha	O curso é pouco procurado, permitindo uma fácil classificação	Outros motivos
Grupo 1	0	0	0	0
Grupo 2	0,0164609	0	0,00411523	0,00411523
Grupo 3	0	0,0267857	0	0,107143
Grupo 4	0,236364	0	0,00909091	0,0818182
Grupo 5	0,0628931	0	0	0,0754717

No grupo 1, o principal motivo de prestar vestibular na UNIOESTE, respondidos por 100% dos acadêmicos foi “O curso pretendido por mim é de reconhecida qualidade”. No o grupo 2, 96,7% dos acadêmicos responderam “ É uma universidade pública, e atende às condições socioeconômicas da família e minhas”. Já 80,35% dos acadêmicos do grupo 3 responderam que “A UNIOESTE possui o curso que desejo cursar”. Para 58,18% dos acadêmicos do grupo 4, o principal motivo de prestar o vestibular na UNIOESTE foi “É uma universidade pública, e atende às condições socioeconômicas da família e minhas” e para 23,63% o motivo é que “A Universidade oferece o curso noturno pretendido, possibilitando-me trabalhar de dia”. Já para 33,33% dos acadêmicos do grupo 5 “A UNIOESTE possui



o curso que desejo cursar” e para 32,07% “É uma universidade pública, e atende às condições socioeconômicas da família e minhas”.

Na tabela 7 são apresentados os centróides para cada grupo para as variáveis: Candidato portador de necessidades especiais, Sexo Feminino e Faixa Etária.

Tabela 7 Centróides dos grupos (conjunto de variáveis 7)

Grupos	Candidato Portador de Necessidades Especiais	Sexo Feminino	Faixa Etária
Grupo 1	0	0,643678	0,747126
Grupo 2	0,0205761	0,748971	0,572016
Grupo 3	0	0,607143	0,892857
Grupo 4	0	0,390909	3
Grupo 5	0	0,465409	0,72956

Em relação a Candidato Portador de Necessidades Especiais, o grupo 2 possui 5 pessoas que representam 2,95% do grupo, sendo 3 acadêmicos com baixa visão, 1 acadêmico com deficiência física e 1 acadêmico que respondeu que possuía necessidade especial mas que quando questionado sobre o tipo da necessidade, informou que não possui necessidades especiais.

Com relação ao sexo o grupo 2 possui 74,89% dos acadêmicos do sexo Feminino, os grupos 1 e 3 também possuem maioria dos acadêmicos do sexo Feminino. Já o grupo 4 possui apenas 39,09% dos acadêmicos do sexo Feminino, sendo que a maioria deste grupo 60,91% são do sexo Masculino. O grupo 5 também possui maioria masculina com 53,46% dos acadêmicos.

Em relação a Faixa Etária observou-se que o grupo 4 tem maior média de idade, sendo que 41,82% dos acadêmicos tem idade acima de 26 anos. Já o grupo 2 tem menor média de idade, sendo que 94,24% dos acadêmicos possuem idade até 20 anos. Nos grupos 1, 3 e 5, os percentuais de acadêmicos com idade até 20 anos são 85,06%, 81,25% e 84,28%, respectivamente.

Apresentam-se na sequencia as principais características de cada grupo. No Grupo 1 os acadêmicos são, em maioria, do sexo feminino, possuem idade até 20 anos, residem no Paraná, frequentaram cursinho preparatório para o vestibular e o principal motivo pelo qual prestaram vestibular na UNIOESTE é que o curso pretendido por eles é de reconhecida qualidade.

No Grupo 2 os acadêmicos são, em maioria, do sexo feminino, possuem



idade até 20 anos (sendo este o grupo mais jovem) e o principal motivo pelo qual prestaram vestibular na UNIOESTE é que esta é uma universidade pública e atende às condições socioeconômicas da família. Além disso, apenas 29,63% dos acadêmicos trabalham, este é o grupo com a menor renda familiar e que tem a maior porcentagem de acadêmicos que estudaram em escola pública.

No Grupo 3 os acadêmicos são, em maioria, do sexo feminino, possuem idade até 20 anos, cursaram o Ensino Fundamental e Médio em escola pública (sendo o segundo grupo com a maior porcentagem de acadêmicos que estudaram em escola pública), o principal motivo pelo qual prestaram vestibular na UNIOESTE é que a universidade possui o curso que desejam cursar e 2,95% possui alguma necessidade especial sendo 3 acadêmicos com baixa visão e 1 acadêmico com deficiência física.

No Grupo 4 os acadêmicos são, em maioria, do sexo masculino, filhos de pais que não concluíram o Ensino Médio, cerca de metade dos acadêmicos do grupo tem idade acima de 26 anos (sendo o grupo menos jovem de todos), possuem casa própria ou financiada, é o grupo com a maior porcentagem de trabalhadores, tem renda familiar inferior a 2 salários mínimos, cursaram Ensino Médio no período noturno e apenas 26,36% frequentaram cursinho. Além disso, a maior parte dos acadêmicos do grupo já havia iniciado algum curso superior anteriormente e, para a maioria, o principal motivo de prestar o vestibular na UNIOESTE é que é uma universidade pública e atende às condições socioeconômicas da família, seguido do fato de que a Universidade oferece o curso noturno pretendido, possibilitando trabalhar durante o dia.

No Grupo 5 os acadêmicos são, em maioria, do sexo masculino, possuem idade até 20 anos, são filhos de pais que possuem graduação e/ou pós graduação, é o grupo com a menor porcentagem de trabalhadores, possuem renda familiar superior a 5 salários mínimos, apenas 6,91% cursou o Ensino Médio em escola pública e foi o grupo com maior porcentagem de acadêmicos que cursaram o Ensino Médio no período diurno. Além disso, para 33,33% dos acadêmicos do grupo, o principal motivo de prestar vestibular na UNIOESTE é que a universidade possui o curso que desejam cursar e para 32,07% é uma universidade pública e atende às condições socioeconômicas da família.

Analisando os grupos obtidos pela aplicação do *K*-médias em relação aos



centros, observa-se que não houve distribuição representativa por centro, nenhum grupo possui maioria absoluta (em percentual) de nenhum centro. Na tabela 8, em cada centro, o maior percentual está representado em negrito, o segundo maior está representado em itálico e o terceiro maior, sublinhado. Para cada grupo, o maior percentual está identificado em vermelho.

Tabela 8 K-médias com 5 centróides e distância Euclidiana

Centros	Grupo 1	Grupo 2	Grupo 3	Grupo 4	Grupo 5
CCET	11,20%	28,00%	24,20%	9,30%	<u>27,30%</u>
CCSA	9,20%	27,00%	9,20%	30,50%	<u>24,10%</u>
CECA	10,40%	43,20%	<u>13,60%</u>	24,80%	8,00%
CCBS	<u>14,70%</u>	38,70%	16,20%	8,30%	<u>22,10%</u>
CCMF	16,30%	33,80%	12,50%	5,00%	32,50%

Analisando a tabela 8 observa-se que o grupo 2 pode ser representado pelo CECA, pois é o grupo no qual o percentual de acadêmicos é o maior, porém, outros centros como CCBS e CCMF também apresentam alto percentual de acadêmicos no grupo. Já o grupo 5 pode ser representado pelo CCMF que possui maior percentual de acadêmicos no grupo, entretanto outros centros como CCSA e CCET também apresentam alto percentual de acadêmicos no grupo. O grupo 4 pode ser representado pelo CCSA que possui o maior percentual de acadêmicos no grupo, contudo, o CECA também apresenta alto percentual de acadêmicos no grupo. O grupo 3 pode ser representado pelo CCET e o grupo 1 pode ser representado pelo CCMF, que possuem maior percentual de acadêmicos no grupo, todavia esses percentuais são baixos.

A tabela 9 apresenta a quantidade de acadêmicos por curso, na divisão dos grupos formados. Em cada curso, o maior percentual está representado em negrito e o segundo maior está representado em itálico.

O grupo 1 não possui maioria de nenhum curso porém possui considerável percentual (mais de 20%) de acadêmicos dos cursos de Fisioterapia, Farmácia e Letras Português/Italiano.

O grupo 2 tem maioria dos acadêmicos dos cursos de Ciências Biológicas Integral e Noturno, Enfermagem, Fisioterapia, Engenharia Agrícola, Farmácia, Ciências Contábeis, Letras Português/Espanhol e Português/Italiano, Pedagogia Matutino e Noturno.



O grupo 3 tem maioria dos acadêmicos do curso de Matemática. Além disso, tem considerável percentual (mais de 20%) de acadêmicos dos cursos de Ciências Biológicas Integral, Fisioterapia, Ciências da Computação, Engenharia Agrícola, Letras Português/Italiano, Letras Português/Espanhol e Letras Português/Inglês.

O grupo 4 tem maioria dos acadêmicos dos cursos de Administração e Ciências Econômicas. Além disso, tem considerável percentual (mais de 20%) de acadêmicos dos cursos de Ciências Biológicas Noturno, Matemática, Ciências Contábeis, Letras Português/Italiano, Letras Português/Espanhol, Pedagogia Matutino e Noturno. Este grupo é formado, em grande maioria, por acadêmicos de cursos noturnos.

O grupo 5 tem maioria dos acadêmicos dos cursos de Odontologia, Ciências da Computação, Engenharia Civil, Medicina e Letras Português/Inglês. Além disso, tem considerável percentual (mais de 20%) de acadêmicos dos cursos de Ciências Biológicas Integral, Fisioterapia, Administração, Ciências Contábeis e Ciências Econômicas.

Tabela 9 Quantidade de acadêmicos por curso em cada grupo

Centro	Curso		G1	G2	G3	G4	G5
CCBS	Ciências Biológicas	I	5,00%	37,50%	30,00%	2,50%	25,00%
CCBS	Ciências Biológicas	N	14,29%	33,34%	11,90%	28,57%	11,90%
CCBS	Enfermagem	I	16,67%	61,90%	4,77%	9,52%	7,14%
CCBS	Fisioterapia	I	20,00%	35,00%	22,50%	0,00%	22,50%
CCBS	Odontologia	I	17,50%	25,00%	12,50%	0,00%	45,00%
CCET	Ciência da Computação	I	15,00%	20,00%	22,50%	2,50%	40,00%
CCET	Engenharia Agrícola	I	9,75%	41,46%	29,27%	12,20%	7,32%
CCET	Engenharia Civil	I	12,50%	20,00%	12,50%	0,00%	55,00%
CCET	Matemática	N	7,50%	30,00%	32,50%	22,50%	7,50%
CCMF	Farmácia	I	20,00%	42,50%	15,00%	5,00%	17,50%
CCMF	Medicina	I	12,50%	25,00%	10,00%	5,00%	47,50%
CCSA	Administração	N	9,81%	27,45%	7,84%	29,41%	25,49%
CCSA	Ciências Contábeis	N	17,50%	35,00%	5,00%	20,00%	22,50%
CCSA	Ciências Econômicas	N	2,00%	20,00%	14,00%	40,00%	24,00%
CECA	Letras Português/Espanhol	M	6,67%	53,33%	20,00%	20,00%	0,00%
CECA	Letras Português/Inglês	M	9,52%	23,81%	23,81%	14,29%	28,57%
CECA	Letras Português/Italiano	M	22,22%	33,34%	22,22%	22,22%	0,00%
CECA	Pedagogia	M	7,50%	50,00%	2,50%	30,00%	10,00%
CECA	Pedagogia	N	12,50%	45,00%	15,00%	27,50%	0,00%



CONSIDERAÇÕES FINAIS

O estudo feito com os dados do questionário socioeconômico dos ingressantes/2012 na Universidade Estadual do Oeste do Paraná – *Campus Cascavel*, utilizando-se a Análise de Agrupamentos, permitiu observar diferentes características nos grupos de acadêmicos formados e identificar a formação destes grupos. Foi possível observar que o grupo 5 é um grupo formado por jovens, dos quais a maioria mora com os pais, possuem renda familiar elevada e estudaram em escola particular, ou seja, é um grupo formado pela “elite” da Universidade. No entanto os acadêmicos pertencentes a este grupo são dos mais variados cursos e não apenas de cursos de maior prestígio.

Já o grupo 4, é formado por acadêmicos com idade superior aos demais grupos, trabalham fora, possuem uma renda familiar inferior em comparação com os demais e seus pais (em sua grande maioria) não concluíram o Ensino Médio. Os acadêmicos deste grupo também são dos mais diversos cursos, inclusive daqueles considerados de maior prestígio.

Assim, a realidade encontrada na Universidade, foi de um avanço na democratização do ensino superior. Esta realidade pode estar relacionada a diversos fatores como a forma de ingresso (incluindo o sistema de cotas), a oferta de bolsas (de Iniciação Científica, Monitoria, etc), entre outros. Diante do exposto, conclui-se que a Universidade pode continuar a desenvolver ações que busquem a democratização do ensino superior.

Além disso, o presente estudo poderá ser utilizado como subsídio para tomada de decisão dentro da Universidade, na elaboração de projetos que se encaixem com os perfis encontrados, garantindo a permanência dos acadêmicos na Universidade.

Entre as muitas ações que poderiam ser executadas, pode-se citar a criação de Restaurantes Universitários, que reduziria o custo de sua alimentação. Outra ação está relacionada à oferta de bolsas (PIBIC, PIBID, PIBITI, Monitoria e Extensão), que poderia ser ampliada, dando oportunidades aos acadêmicos de se manterem, sem a necessidade de trabalhar e ainda privilegiando sua formação.



REFERÊNCIAS

CARVALHO, CRISTINA HELENA ALMEIDA . **O PROUNI no governo Lula e o jogo político em torno do acesso ao ensino superior.** *Educ. Soc.*, Campinas, vol. 27, n. 96 - Especial, Out. 2006

DINIZ, C. A. R.; LOUZADA NETO, F. **Data mining: uma introdução.** São Paulo: ABE, 2000.

FAYYAD, U. M.; PIATETSKY-SHAPIO, G.; SMYTH, P.; UTHRUSAMY, R. **Advances in Knowledge Discovery & Data Mining.** California: AAAI/MIT, 1996.

HAIR JR, J.F.; ANDERSON, R.E.; TATHAM, R.L.; BLACK, W.C. **Análise Multivariada de Dados.** Tradução de: SANTANNA, A. S.; CHAVES NETO, A. Porto Alegre: Bookman, 2005.

LAGE, Flávia Temporim de Oliveira; **Democratização do Ensino Superior no Brasil.** Rio de Janeiro, Mar. 2007.

MARTINHAGO, S. **Descoberta de conhecimento sobre o processo seletivo da UFPR.** Dissertação de Mestrado. UFPR, Curitiba, 2005.

MOREIRA, Plínio Cavalcanti ET AL. **Quem quer ser professor de matemática?.** Zetetiké – FE/Unicamp – v. 20, n. 37 – jan-jun 2012.

OLIVEIRA, Ionara Stéfani Viana de. **Os Determinantes do desempenho acadêmico do corpo docente no Ensino Superior: Evidências a partir da Universidade Federal da Paraíba.** Mestrado em Economia. UFPB, 2011.

SOUSA, Paulo Germano; OLIVEIRA Neto, Raimundo Cardoso; NASCIMENTO, Fernando Ferraz do Nascimento. **Análise socioeconômica dos alunos da UFPI, campus Ministro Petrônio Portela.** Revista Simed, 2009.

SPÍNDOLA, Thelma; MARTINS, Elizabeth Rose da Costa; FRANCISCO, Márcio Tadeu Ribeiro. **Enfermagem como opção: perfil de graduandos de duas instituições de ensino.** Rev. bras. enferm., Abr. 2008, vol.61, no.2, p.164-169.

TAN, Pang – Ning; STEINBACH, Michael; KUMAR, Vipin. **Introdução ao DATA MINING Mineração de Dados.** Rio de Janeiro: Editora Ciência Moderna Ltda, 2009.

UEMURA, E. **Reflexões sobre o planejamento da Universidade,** São Paulo: Revista Science in Health, 2010.

UNIOESTE. **A Universidade Estadual do Oeste do Paraná - Campus de Cascavel;** [acesso 15 set. 2012]. Disponível em: http://www.cascavel.unioeste.br/index.php?option=com_content&view=article&id=14&Itemid=19

