

Título do Projeto:
Tecnologias de Mineração de Dados aplicadas à Educação a Distância

Linha 02: Ferramentas Tecnológicas para Educação a Distância

Orientador: Domingos Sávio Pereira Salazar

Programa de Pós-Graduação em Tecnologia e Gestão em Educação a Distância

Unidade Acadêmica de Educação a Distância e Tecnologia - Universidade Federal Rural de Pernambuco

Resumo e Justificativa:

Discentes da modalidade de Educação a Distância possuem características interessantes. Em geral, a quantidade e a dispersão geográfica de alunos vinculados a um curso de EAD é bem maior em comparação ao presencial, fatos que aumentam a relevância estatística nas amostras. Neste projeto, dados cadastrais dos discentes, como idade, endereço, e nota do Enem, por exemplo, serão combinados aos dados do município (polo) em questão para gerar um modelo preditivo capaz de calcular o potencial de sucesso de cada discente e de cada polo em um dado curso EAD.

Objetivo:

Utilizar técnicas de mineração de dados e aprendizado de máquina para detectar padrões em base de dados de discentes de educação a distância para detectar o sucesso de investimentos e abertura de cursos em diferentes polos (municípios) de apoio presencial.

Objetivos Específicos:

1. Estudar técnicas de mineração de dados e aprendizado de máquina para problemas de classificação binária e regressão com base na metodologia CRISP-DM.
2. Realizar aquisição e preparação de base de dados integradas utilizando fontes de dados disponíveis (como UAB, CAPES, Sig@, IBGE).
3. Testar diferentes metodologias de Mineração de Dados na base de dados gerada no item anterior com objetivo de estimar o a melhor alternativa possível de alocação de recursos em diferentes polos de apoio presencial.
4. Análise de desempenho da solução sob diferentes métricas de avaliação de desempenho em forma de relatório para divulgação.

Resultados Esperados

Dentre outras aplicações, o modelo pode ser utilizado para calcular qual a melhor alocação de cursos (e quantidade de vagas) em diferentes polos de apoio presencial nas ofertas da Universidade Aberta do Brasil a fim de atingir o maior sucesso possível para um valor limitado de recursos.

Perfil do Candidato:

Graduado em Estatística, Ciências da Computação ou áreas afins.

Habilidades Essenciais:

- Uso avançado de ferramentas de MS Excel (ou Open Office).
- Noções básicas ou alguma experiência com banco de dados e modelos de classificação binária e regressão.
- Experiência com pelo menos uma linguagem de programação.

Principais Responsabilidades no Projeto:

- Realizar entendimento e definição do problema a partir de leitura de publicações científicas.
- Realizar aquisição, preparação e enriquecimento de base de dados distintas (Mec, CAPES, INEP, IBGE, etc.) a partir do entendimento do problema.
- Utilizar diferentes técnicas de mineração de dados presentes na literatura científica para resolver o problema.
- Realizar análise de desempenho do modelo de acordo com métricas de avaliação especificadas nas etapas anteriores; escrever relatório com resultados e artigo científico em colaboração; apresentar os resultados em eventos da área.