# UNIVERSIDADE DE SÃO PAULO - LESTE

**PROGRAMA DE PÓS-GRADUAÇÃO EM SISTEMAS DE INFORMAÇÃO SYLABUS DA DISCIPLINA: MINERAÇÃO DE PROCESSOS**

1. **DADOS INFORMATIVOS**
   1. Semestre : II - 2022
   2. Horas semanais : 4 horas semanais (60h – 8 créditos)
   3. Docentes : Profa. Dra. Sarajane Marques Peres ([sarajane@usp.br](mailto:sarajane@usp.br))

Prof. Dr. Marcelo Fantinato ([m.fantinato@usp.br](mailto:m.fantinato@usp.br))

* 1. Código : SIN5025
  2. Sala : Sala de aula virtual usando o Sistema e-Disciplinas

# DESCRIÇÃO

**Objetivos**:

* Capacitar o aluno a: identificar contextos de aplicação para mineração de processos; identificar as técnicas, os métodos e as ferramentas mais adequadas para tratar tipos específicos de mineração de processos; e avaliar a qualidade de resultados obtidos a partir da aplicação de mineração de processos.
* Explorar os tipos de mineração de processo em contextos teóricos e aplicados, considerando problemas sintéticos e do mundo real.
* Fornecer conhecimentos de base teórica para capacitar o aluno a trabalhar em prol do desenvolvimento científico da área de mineração de processos.
* Proporcionar um ambiente de prática para implementação e aplicação de técnicas de mineração de processos e para treinamento no uso de ferramentas de prateleira.

# Justificativa:

A análise e a melhoria de processos de negócio dependem de modelos de processo que formalizam as caraterísticas dos processos e evidenciam suas virtudes e fraquezas. Porém, modelos de processo nem sempre estão disponíveis nas organizações, mostrando a falta de conhecimento sobre como seus processos são executados. A mineração de processos pode lidar com essa questão prática da vida organizacional por meio da análise de logs de eventos produzidos pelos sistemas de informação. Nesse contexto, as áreas de estatística, inteligência artificial, aprendizado de máquina e reconhecimento de padrões são essenciais, pois permitem extrair informações úteis de logs de eventos para construir modelos de processos bem como para verificar e melhorar tais modelos. Mineração de processos é uma área de pesquisa e aplicação relativamente nova que une conceitos de gestão de processos e de mineração de dados. Embora sejam empregados conceitos sedimentados de administração, estatística e computação, novos problemas são trazidos devido a necessidade de análise de “eventos”, objetos de análise diferente dos “dados” tradicionalmente analisados em mineração de dados. O tratamento desses novos problemas propicia o desenvolvimento de pesquisas que possam levar à evolução de métodos, técnicas e algoritmos clássicos de análise de dados. Mineração de processos apresenta alto impacto para a indústria de sistemas de informação e de tecnologia da informação, como mencionado pela consultoria Gartner em 2018. O “Process Mining Manifesto” (de 2012), assinado por 77 pesquisadores de 53 instituições – principalmente europeias – e chancelado pelo Institute of Electrical and Electronic Engineers (IEEE), selou a importância dessa área ao estabelecer uma terminologia básica e criar uma força tarefa para “aumentar a maturidade no desenvolvimento de ferramentas para melhorar o (re)projeto, controle e apoio de processos de negócio operacionais”.

# Conteúdo:

1. Contextualização e motivação: visão geral, ciência de dados X ciência de processo, exemplos motivacionais de casos de uso do mundo real e fictícios, problemas e tendências.
2. Preliminares: processos de negócios e mineração de dados; visão geral sobre gestão de processos de negócio (BPM

– Business Process Management) – notações para modelagem de processos; visão geral sobre mineração de dados

(DM – Data Mining) – tarefas e técnicas básicas; relacionamento entre mineração de processos e BPM, mineração de processos e DM, mineração de processos e outros conceitos (business intelligence, big data etc.).

1. Processos do tipo “lasanha” e do tipo “espaguete”: caracterização, casos de uso, ciclo de vida, aplicações.
2. Logs de eventos: fonte de dados, formalização, questões de qualidade de dados.
3. Descoberta de processos: conceitos básicos, algoritmo Alpha, representação, ruído e incompletude, critérios de qualidade, mineração heurística, mineração de processos usando computação evolutiva, mineração indutiva, outras técnicas de descoberta de processos; avaliações, casos de uso, problemas e aplicações.
4. Verificação de conformidade e aprimoramento de processos: conceitos básicos, alinhamento e auditoria, repetição e comparação, reparo e avaliação; avaliação, casos de uso, problemas e aplicações.
5. Ferramentas de apoio: ProM, Apromore, Disco etc.

# 3. PROGRAMAÇÃO

|  |  |  |  |
| --- | --- | --- | --- |
| **Aula** | **Data** | **Conteúdo / atividade prevista (parte I)** | **Conteúdo / atividade prevista (parte II)** |
| **1** | 18 de agosto | Apresentação da disciplina/alunas e alunos | Visão geral de mineração de processos |
| **2** | 25 de agosto | Apresentação da disciplina/alunas e alunos | Visão geral de mineração de processos |
| **3** | 1 de setembro | Modelos de processo | Modelos de processo |
| **4** | 15 de setembro | Modelos de processo | Dados / logs |
| **5** | 22 de setembro | Dados / logs | Descoberta de modelos de processo |
| **6** | 29 de setembro | Descoberta de modelos de processo | Descoberta de modelos de processo |
| **7** | 6 de outubro | Descoberta de modelos de processo | Verificação de conformidade |
| **8** | 13 de outubro | Verificação de conformidade | Verificação de conformidade |
| **9** | 20 de outubro | Apresentação da Everflow (a confirmar) | Tópicos em mineração de processos (suporte e melhoria) |
| **10** | 27 de outubro | Apresentação da Apromore/PM4PY (a confirmar) | Tópicos em mineração de processos (suporte e melhoria) |
| **11** | 3 de novembro | **Apresentação de artigos (1, 2, 3)** | |
| **12** | 10 de novembro | **Apresentação de artigos (4, 5, 6)** | |
| **13** | 17 de novembro | Apresentação da Upflux (a confirmar) | Aula professor convidado |
| **14** | 24 de novembro | Apresentação da Celonis (a confirmar) | Tópicos em mineração de processos (suporte e melhoria) |
| **15** | 1 de dezembro | **Apresentação de projetos (1, 2, 3, 4, 5, 6)** | |

1. **BIBLIOGRAFIA GERAL**
   * [1] Wil van der Aalst, Process Mining: Data Science in Action, Springer Publishing Company, Inc., 2nd edition, 2016.
   * [2] Ana Rocío Cárdenas Maita, Lucas Corrêa Martins, Carlos Ramón López Paz, Laura Rafferty, Patrick C. K. Hung, Sarajane Marques Peres, Marcelo Fantinato, A systematic mapping study of process mining, Enterprise Information Systems, vol. 12, n. 5, pp. 505-549, 2018.
   * [3] Wil van der Aalst et al., Process mining manifesto, 7th International Workshop on Business Process Intelligence (BPI 2011), Springer, Berlin, Heidelberg, pp. 169-194, 2012.
   * [4] Andrea Burattin, Process Mining Techniques in Business Environments – Theoretical Aspects, Algorithms, Techniques and Open Challenges in Process Mining, series Lecture Notes in Business Information Processing, vol. 207, Springer, 2015.
   * [5] W. M. P. van der Aalst, H. A. Reijers, A. J. M. M. Weijters, B. F. van Dongen, A. K. Alves de Medeiros, M. Song and H.

M. W. Verbeek, Business process mining: An industrial application, Information Systems, vol. 32, n. 5, pp. 713-732, 2007.

* + [6] Marlon Dumas, Marcello La Rosa, Jan Mendling, Hajo A. Reijers, Fundamentals of Business Process Management, 2nd edition, Springer, 2018.
  + [7] Mathias Weske, Business Process Management: Concepts, Languages, Architectures. Springer, Berlin, Heidelberg, 2007.
  + [8] I. H. Witten, E. Frank, M. A. Hall, C. J. Pal, Data Mining: Practical Machine Learning Tools and Techniques, 4th edition, The Morgan Kaufmann, 2016.
  + [9] J. Han, M. Kamber, J. Pei, Data Mining: Concepts and Techniques, 3rd edition, The Morgan Kaufmann Publishers, 2011.
  + [10] C. C. Aggarwal, Data Mining: The Textbook, Springer, 2015.

# METODOLOGIA

* + Exposição em aula pelos docentes dos temas do curso, uso de material audiovisual e palestras de convidados.
  + Realização de avaliação prática (projeto final e apresentação de artigos).
  + Utilização do sistema de gestão de cursos e-Disciplinas para as aulas online (síncronas), disponibilização de material e para comunicação de recados gerais para a turma.

# EXERCÍCIOS E TRABALHOS

* + Realização de exercícios, apresentações e projetos em **grupos de 3 alunos**. O projeto e a apresentação de artigos serão feitos com a mesma formação de grupo.
  + **Projeto:** implementação e avaliação de uma estratégia para resolução de problemas no contexto de mineração de processos.
    - Cada grupo **deverá escolher tratar uma das seguintes tarefas**:
      * Implementar e avaliar uma estratégia para execução e análise de *trace clustering*.
      * Implementar e avaliar uma estratégia para detecção e reparação de traces ruidosos/anômalos.
      * Implementar e avaliar uma estratégia para detecção de *concept drifts*.
      * ~~Implementar e avaliar uma estratégia para predição de tempo de conclusão do caso.~~
      * ~~Implementar e avaliar uma estratégia para recomendação de recurso.~~
    - **Será fornecido um modelo de processo que deverá ser usado para geração de logs sintéticos. O grupo poderá inserir complexidade no modelo fornecido, porém não poderá simplificá-lo.**
    - Cada grupo deverá manipular esse log de eventos sintético de forma a inserir **uma (ou mais, se adequado)** das seguintes características:
      * Ruído / anomalias.
      * *Concept drifts.*
      * ~~Informações referentes a recursos.~~
      * ~~Informações referentes a custo.~~
      * Informações referentes a tempo.
    - Cada grupo deverá testar a sua abordagem em um log de eventos do mundo real, usando:
      * Log de eventos referentes aos desafios do BPI/ICPM (2011 a 2022).
      * Log de eventos referentes a movimentação de processos jurídicos (Process Mining @ USP).
      * Log de eventos referentes a gestão de incidentes (Process Mining @ USP).
      * Log de eventos referentes a processos de pedidos da Anvisa.
    - A descrição da estratégia implementada assim como da sua aplicação sobre os logs de eventos deverá ser organizada em um relatório que deverá estar formatado seguindo o template adotado por periódicos IEEE (template IEEEtran).
    - O relatório deverá ter no mínimo 6 páginas.

# Data de entrega final: 1 de dezembro até meio-dia.

* + - Entrega do relatório em formato PDF via e-Disciplinas.
    - Artefatos de codificação também deverão ser depositados no e-Disciplinas, acompanhados de um log de eventos que permita a execução dos códigos e de um arquivo *readme* com instruções para execução.
    - Apenas um aluno do grupo precisará depositar o relatório e os artefatos no sistema.
    - A apresentação do trabalho deverá ter duração de até **20 minutos**. Todos os membros do grupo deverão apresentar. Se algum membro não apresentar, ele ficará com 30% da nota recebida pelo grupo, se os demais membros confirmarem que ele participou da elaboração do trabalho.
  + **Apresentação de artigo:** apresentação de artigos em aula online, para todos os participantes.
    - O artigo a ser apresentado pelo grupo será sorteado a partir de um conjunto de artigos determinado pelos professores da disciplina.
    - A apresentação deverá ter duração de até **40 minutos**. Todos os membros do grupo deverão apresentar. Se algum membro não apresentar, ele ficará com 30% da nota recebida pelo grupo, se os demais membros confirmarem que ele participou da elaboração do trabalho.
    - O grupo não deverá se ater apenas ao artigo sorteado, mas também explorar artigos correlatos ou artigos anteriores dos mesmos autores, desenvolvidos para o mesmo problema.
    - O grupo deverá preparar um conjunto de slides para suportar sua apresentação. Esse conjunto de slides deverá ser depositado no e-Disciplinas, e ficará disponível para todos os participantes.

# Data de apresentação: 3 ou 10 de novembro – sob sorteio.

* + - Entrega dos slides via e-Disciplinas no dia da apresentação, até meio-dia.
    - Apenas um aluno do grupo precisará depositar os slides no sistema.
    - A nota atribuída aos membros do grupo pode ser diferente, sendo dependente do desempenho de cada aluno na apresentação.

# AVALIAÇÃO DO APRENDIZADO

A avaliação do aprendizado será feita utilizando as seguintes diretrizes:

# O aluno deve alcançar a frequência mínima de 75% das aulas.

* + Avaliação:
* Os seguintes itens serão avaliados:
  + Projeto (P)
  + Apresentação de artigo (AA)
* A Média Final (MF) do aluno obedecerá à seguinte regra:

# MF = (7\*P + 3\*AA) / 10

* O mapeamento da MF para conceitos se dará da seguinte forma:

|  |  |
| --- | --- |
| **Média Final (MF)** | **Conceito** |
| 8,0 ≤ MF ≤ 10,0 | A |
| 6,5 ≤ MF < 8,0 | B |
| 5,0 ≤ MF < 6,5 | C |
| 0,0 ≤ MF < 5,0 | R |

# POLÍTICAS

* + A programação de aulas é preliminar e está sujeita a mudanças.
  + A programação de entregas e apresentações de trabalhos será alterada apenas por motivos consequentes de ações ou determinações acadêmicas.
  + Não serão admitidos atrasos na entrega do trabalho, sendo que o aluno que não entregar receberá nota 0 na respectiva avaliação.
  + Não existe abono de faltas em nenhuma situação (nem mesmo por motivo de saúde). Procure “guardar” os 25% de possibilidades de falta para quando REALMENTE for necessário faltar.
    - Eventuais problemas referentes à disciplina deverão ser tratados junto aos professores, logo após o término das aulas ou em reuniões agendas *ad hoc*.