

<div><div>PREDICTION TASK</div><div></div></div> <div>Type of task? Entity on which predictions are made? Possible outcomes? Wait time before observation?</div> <div>1 - Preditivo</div>	<div><div>DECISIONS</div><div></div></div> <div>How are predictions turned into proposed value for the end-user? Mention parameters of the process / application that does that.</div> <div>1 - Ao encontrar os motivos, realizar campanhas preventivas focadas em diminuir a inadimplência dos determinados motivos</div>	<div><div>VALUE PROPOSITION</div><div></div></div> <div>Who is the end-user? What are their objectives? How will they benefit from the ML system? Mention workflow/interfaces.</div> <div>1 - Realizar estudo/descoberta de associações em base de clientes de uma empresa de previdência privada para descobrir/relacionar os motivos pelos quais os clientes ficam inadimplentes com suas contribuições;</div> <div>2 - Prever taxa de inadimplência do ano subsequente.</div>	<div><div>DATA COLLECTION</div><div></div></div> <div>Strategy for initial train set & continuous update. Mention collection rate, holdout on production entities, cost/constraints to observe outcomes.</div> <div>1 - Histórico de clientes inadimplentes do último ano da base Sinqia/Previc;</div> <div>2 - Acontecimentos importantes a nível BR como, copa do mundo, pandemia, política.</div>	<div><div>DATA SOURCES</div><div></div></div> <div>Where can we get (raw) information on entities and observed outcomes? Mention database tables, API methods, websites to scrape, etc.</div> <div>1 - Base de dados do sistema de previdência privada da Sinqia;</div> <div>2 - Dados abertos do órgão regulador Previc;</div> <div>3 - Correlações com variáveis macroeconômicas como taxa de endividamento e desemprego, além de correlações com a própria base como entrada na aposentadoria por idade/invalidez/morte.</div>		
<div><div>IMPACT SIMULATION</div><div></div></div> <div>Can models be deployed? Which test data to assess performance? Cost/gain values for (in)correct decisions? <u>Fairness constraint</u>?</div> <div>1 - Treinar o modelo em cada motivo; 2 - Realizar testes com dados reais; 3 - Conversar com a área mostrando os resultados para ver se faz sentido.</div>	<div><div>MAKING PREDICTIONS</div><div></div></div> <div>When do we make real-time / batch pred.? Time available for this + featurization + post-processing? Compute target?</div> <div>1 - Predição de inadimplência com base no último trimestre.</div>	<div><div>MONITORING</div><div></div></div> <div>Metrics to quantify value creation and measure the ML system's impact in production (on end-users and business)?</div> <div>1 - Acompanhamento da taxa de clientes inadimplentes a cada trimestre.</div> <div>2 - Comparação da taxa de inadimplência entre os períodos.</div>			<div><div>BUILDING MODELS</div><div></div></div> <div>How many prod models are needed? When would we update? Time available for this (including featurization and analysis)?</div> <div>1 - EDA</div> <div>2 - Descoberta de Associações.</div> <div>3 - Supervisionado: Classificação</div> <div>4 - Não Supervisionado - Clusterização</div>	<div><div>FEATURES</div><div></div></div> <div>Input representations available at prediction time, extracted from raw data sources.</div> <div>Target: CPF do participante inadimplente, pois precisamos analisar o motivo de cada participante para clusterizar os principais motivos pela inadimplência</div> <div>-Salário; -Dependentes; -Profissão; -Tempo de contribuição; -Valor total contribuído; -Tempo inadimplência;</div> <div>Variáveis: -Datas comemorativas; -Pagamento de impostos (IPVA,IPTU); -Endividamento; -Morte/Invalidez/Tempo.</div>