

## 1.0 Objeto

1.1. Consulta Pública para Contratação de Produto para Construção de Soluções de Inteligência Artificial por meio de Automated machine learning (AutoML)

## 2.0 Especificação do objeto a ser contratado

2.1. Produto para Construção de Soluções de Inteligência Artificial por meio de Automated machine learning (AutoML).

### 2.1.1. Introdução

2.1.1.1. Dadas as crescentes demandas por soluções de inteligência artificial - IA, identificou-se a necessidade de contratação de um produto que contenha um conjunto (suíte) de ferramentas que possibilitem ao SERPRO construir soluções de IA com alta produtividade e em larga escala e, além disso, ofertar para seus clientes uma plataforma tecnológica que possibilite aos seus especialistas a criação de suas próprias soluções.

Tendo isso em vista, este documento apresenta uma visão inicial, a ser complementada após consulta pública, de requisitos e características técnicas que serão consideradas para contratação de um produto para construção de soluções de inteligência artificial por meio de automated machine learning (AutoML).

2.1.1.1.1. Automated machine learning (**AutoML**) é o processo de automatizar de ponta a ponta a aplicação de aprendizado de máquina a problemas do mundo real. Em uma aplicação típica de aprendizado de máquina, são aplicados métodos e funcionalidades apropriados de **pré-processamento de dados, engenharia, extração e seleção** que tornam o conjunto de dados passível de aprendizado de máquina. Essas etapas de **pré-processamento dos dados, seleção de algoritmos e otimização dos hiperparâmetros** visam maximizar o desempenho preditivo do modelo final de aprendizado de máquina.

### 2.1.2. Requisitos Técnicos

2.1.2.1. O produto para construção de soluções de inteligência artificial por meio de AutoML deverá possibilitar a aplicação de várias técnicas e métodos de análise de dados e inteligência artificial pelos próprios usuários por meio de recursos visuais.

2.1.2.2. Deve permitir no mínimo a realização de:

2.1.2.2.1. Preparação e qualidade de dados:

2.1.2.2.1.1. Permitir o registro e acompanhamento de todas as etapas do pipeline de preparo e transformação dos dados, de preferência de maneira visual (em fluxo);

2.1.2.2.1.2. Integração com gerenciadores de configuração, como Git, para manutenção de scripts e fluxos de transformação e preparo de dados;

2.1.2.2.1.3. Possuir catálogo de dados ou se integrar com um deles. Permitir catalogação de dados e metadados, como: Formato, relacionamentos, propriedade e direitos e localização;

2.1.2.2.1.4. Possuir funcionalidade para análise de qualidade de dados, como: Data profiling (examinar, analisar e rever os dados para coletar estatísticas sobre a qualidade e "higienização" do conjunto de dados), e visualização gráfica com possível "drill down" dos dados, além de identificação e remoção de "outliers" e nulos, além de normalização ou padronização de colunas;

2.1.2.2.1.5. Possuir mecanismo para amostragem e particionamento de dados, como: amostragem aleatória, estratificada, bootstrap, além de separação aleatória de dados em treino e teste;

2.1.2.2.1.6. Possuir mecanismo para aplicar marca d'agua em datasets (Rastreabilidade);

2.1.2.2.1.7. Possuir funcionalidade de descaracterização (ou mascaramento) de dados sensíveis;

2.1.2.2.1.8. Possuir recursos para redução de variáveis ou dimensionalidade;

2.1.2.2.1.9. Possibilitar a geração de gráficos de análise da qualidade de dados: Dispersão (2D e 3D), Histogramas, Correlação de variáveis, Box Plot, Treemap, Mapa de calor;

2.1.2.2.2. Processamento de linguagem natural (PLN) e Análise textual.

2.1.2.2.2.1 O processamento de texto deve suportar funcionalidades/algoritmos;

2.1.2.2.2.2. Remoção de stopwords;

2.1.2.2.2.3. Remoção de tags XML;

2.1.2.2.2.4. Transliteração para remoção de acentuação;

2.1.2.2.2.5. Expansão de contrações (ex: no idioma inglês I'd vira I would);

2.1.2.2.2.6. Remoção de caracteres especiais;

2.1.2.2.2.7. Stemização;

2.1.2.2.2.8. Normalização morfológica (Substituir vários termos por um que tenha um significado amplo (termo representativo));

2.1.2.2.2.9. Lematização;

2.1.2.2.2.10. Segmentação de texto em sentenças e palavras;

2.1.2.2.2.11. Estatísticas de texto (número de parágrafos, sentenças, palavras, caracteres, etc);

2.1.2.2.2.12. Suporte para análise sintática de sentenças.

2.1.2.2.2.13. N-gramas;

2.1.2.2.2.14. Índices de documentos;

2.1.2.2.2.15. Detecção de sequências de palavras de valor semântico único (chunking);

2.1.2.2.2.16. Vetores de palavras (One-Hot-Encoding, CBOW, Skip-Gram, GloVe, fastText, Word2Vec);

2.1.2.2.2.17. Disponibilizar Vetores de Palavras (Word Embeddings) treinados em português;

2.1.2.2.2.18. Bag of words binário;

2.1.2.2.2.19. Bag of words TF;

2.1.2.2.2.20. Bag of words TF-IDF;

2.1.2.2.2.21. LDA (Latent Dirichlet Allocation) - distribuição de tópicos por documento;

2.1.2.2.2.22. LSA (Latent Semantic Analysis) - matriz de documentos;

2.1.2.2.2.23. NMF (Non-negative Matrix Factorization) - matriz de documentos;

2.1.2.2.2.24. Paragraph2Vec (a.k.a. doc2vec).

2.1.2.2.2.25. Recuperação de informação (busca textual) - Indexação de documentos de texto e interface para busca;

2.1.2.2.2.26. Suporte ao idioma português brasileiro (pt-br);

2.1.2.2.2.27. Ranqueamento de termos (ex: por ganho de informação) - Exemplo: ranqueamento dos termos baseado em entropia e ganho de informação;

2.1.2.2.2.28. Suporte ao idioma inglês (en);

2.1.2.2.2.29. Suporte ao idioma espanhol (es);

2.1.2.2.2.30. Tradução automática (linguagens de referência: pt-br, en, es);

2.1.2.2.2.31. Sumarização automática;

2.1.2.2.2.32. Part-of-speech (POS) tagging;

2.1.2.2.2.33. Named entity recognition (NER);

2.1.2.2.2.34. Representação em lógica de primeira ordem. Redução de sentenças em texto a um conjunto de predicados lógicos;

2.1.2.2.2.35. Geração automática de taxonomias e ontologias. Extração de conceitos de texto e seus relacionamentos visando a construção de taxonomias e ontologias (OWL);

2.1.2.2.2.36. Parsing (geração de árvore gramatical);

2.1.2.2.2.37. Resolução de correferência - Encontrar as expressões que dizem respeito à mesma entidade no texto; Por exemplo, na sentença "Jorge está cansado, já que ele trabalhou muito", a expressão "ele" se refere a "Jorge";

2.1.2.2.2.38. Aprendizado sequência a sequência (Seq2Seq) - A aprendizagem sequência a sequência (Seq2Seq) refere-se a modelos para converter sequências de um domínio (por exemplo, frases em inglês) para sequências em outro domínio (por exemplo, as mesmas frases traduzidas para francês);

2.1.2.2.2.39. Baseada em regras - Contagem de frequência de palavras reconhecidamente positivas e negativas;

2.1.2.2.2.40. Baseada em análise léxica - Baseada na polaridade de termos e sentenças no texto.  
Baseada em modelo estatístico - Avaliações de sentimentos segmentados por aspecto, com polaridade definida

como uma distribuição probabilística de avaliações por aspecto;

2.1.2.2.2.41. Baseada em modelo de aprendizado de máquina - Uso de modelos de classificação, agrupamento, descoberta de tópicos, dentre outros, para análise de sentimento;

2.1.2.2.2.42. Para avaliação sob mais de um aspecto - Mais de um aspecto sendo avaliado. Exemplo: em um atendimento de telemarketing, pode-se inferir a avaliação sob aspectos de tempo de atendimento, qualidade da solução, etc;

2.1.2.2.2.43. Para avaliação de polaridade - Avaliação global positiva ou negativa;

2.1.2.2.2.44. Para avaliação baseada em "estrelas" - Sentimento é inferido em gradação por "estrelas";

2.1.2.2.2.45. Modelagem de tópicos por meio dos modelos LDA, LSA, NMF;

2.1.2.2.2.46. Classificação de texto;

2.1.2.2.2.47. Agrupamento de texto;

2.1.2.2.2.48. Seleção de features;

2.1.2.2.2.49. Detecção de anomalias;

2.1.2.2.2.50. Descoberta de regras de associação;

2.1.2.2.2.51. Possuir ferramenta de comparação de textos (similaridade);

2.1.2.2.2.52. Pré-processamento de texto: suportar as funcionalidades/algoritmos;

2.1.2.2.2.53. Representação de texto: suportar as funcionalidades/algoritmos;

2.1.2.2.2.54. Tarefas de NLP em geral (sintaxe, semântica, recuperação de informação, etc): suportar as funcionalidades/algoritmos;

2.1.2.2.2.55. Análise de sentimento;

2.1.2.2.2.56. Mineração de texto (transversal): suportar as funcionalidades/algoritmos;

2.1.2.2.3. Análises preditivas

2.1.2.2.3.1. Possuir uma biblioteca com modelos de Regressão linear: Regressão Simples; Regressão Múltipla; Regressão Não-Linear;

2.1.2.2.3.2. Possuir uma biblioteca com modelos de Regressão logística: Regressão Binária, Regressão Multinomial, Regressão Ordinal;

2.1.2.2.3.3. Permitir análise de regressão com seleções do tipo stepwise: Forward e Backward;

2.1.2.2.3.4. Possuir ferramenta para testar modelos preditivos com no mínimo cross-validation, análise sensitiva e teste A/B;

2.1.2.2.3.5. Possuir suporte a Naive Bayes, Kernel Naive Bayes e Redes Bayesianas;

2.1.2.2.3.6. Possui mecanismos de classificação, árvores de decisão e árvores de regressão (CART) como ID3, C4.5, Chi-square Adjusted Interaction Detection [CHAID], Multivariate Adaptive Regression Splines [MARS], Random forests e Gradient boosting. Crescimento/poda de árvores;

2.1.2.2.3.7. Possuir técnicas de discretização - Geração de modelos de dois estágios, com a finalidade de estimar uma variável intervalar baseada em classes de uma determinada variável categórica;

2.1.2.2.3.8. Possuir as funcionalidades de Avaliação: Curvas "lift" para avaliação simples e de múltiplos modelos, Curvas ROC, Curvas de lucro (baseadas em uma matriz de custo/benefício) e Matriz de confusão;

2.1.2.2.3.9. Suportar algoritmos de Expectation-Maximization (EM) e Agglomerative Hierarchical Clustering;

2.1.2.2.3.10. Possuir suporte a métodos de otimização automatizada de parâmetros (Grid Search, Randomized Search);

2.1.2.2.3.11. Possuir recursos para avaliação do desempenho do treinamento de modelos de regressão incluindo as métricas: Mean Absolute Error (MAE), Mean Square Error (MSE), Root Mean Square Error (RMSE) e Coeficiente de Determinação (R2);

2.1.2.2.3.12. Possuir recursos para avaliação do desempenho do treinamento de modelos de classificação incluindo as métricas: Acurácia, Precisão, Sensibilidade, F-1 (F-measure), ROC Curve e Matriz de Confusão, R2;

2.1.2.2.3.13. Possuir recursos para avaliação da qualidade de agrupamentos (clusters) incluindo as métricas: Homogeneity, Completeness, V-measure e Silhouette Coefficient;

2.1.2.2.3.14. Possibilitar o uso da técnica Regressão Linear (Simples, Múltipla) via interface gráfica da própria ferramenta;

2.1.2.2.3.15. Possibilitar o uso da técnica Regressão Não Linear (Simples, Múltipla) via interface gráfica da própria ferramenta;

2.1.2.2.3.16. Possibilitar o uso da técnica Regressão Logística (Binária, Multinomial, Ordinária) via interface gráfica da própria ferramenta;

2.1.2.2.3.17. Possibilitar o uso da técnica Modelos Lineares Generalizados - MLG's via interface gráfica da própria ferramenta;

2.1.2.2.3.18. Possibilitar o uso da técnica K Vizinhos mais próximos - KNN via interface gráfica da própria ferramenta;

2.1.2.2.3.19. Possibilitar o uso da técnica Árvore de decisão (ID3, C.4.5, C.5.0, CART) via interface gráfica da própria ferramenta;

2.1.2.2.3.20. Possibilitar o uso da técnica Floresta Aleatória - Random Forest via interface gráfica da própria ferramenta;

2.1.2.2.3.21. Possibilitar o uso de técnicas de Agrupamento - Clustering (Hierárquico, Affinity Propagation, K-Means, DBScan, Mean Shift ) via interface gráfica da própria ferramenta;

2.1.2.2.3.22. Possibilitar o uso da técnica Naive Bayes (Gaussian, Multinomial, Bernoulli) via interface gráfica da própria ferramenta;

2.1.2.2.3.23. Possibilitar o uso da técnica SVM (Classificação, Regressão) via interface gráfica da própria ferramenta;

2.1.2.2.3.24. Possibilitar o uso da técnica Redes Neurais (MLP, Adaline) via interface gráfica da própria ferramenta;

2.1.2.2.3.25. Possibilitar o uso da técnica Redes Neurais Profundas (CNN, RNN) via interface gráfica da própria ferramenta;

2.1.2.2.3.26. Possibilitar a descoberta de regras de associação em bases de dados através das técnicas Apriori, Eclat e FP-growth, via interface gráfica da própria ferramenta;

2.1.2.2.3.27. Possibilitar o uso da técnica Filtragem Colaborativa para construção de recomendadores de conteúdo via interface gráfica da própria ferramenta;

2.1.2.2.3.28. Possibilitar o uso da técnica Extremely Randomized Trees (ExtraTreesClassifier) via interface gráfica da própria ferramenta;

2.1.2.2.3.29. Possibilitar o uso da técnica Gradient Boosting (XGBoost) via interface gráfica da própria ferramenta;

2.1.2.2.3.30. Possibilitar o uso da técnica AdaBoost via interface gráfica da própria ferramenta;

2.1.2.2.3.31. Análise de redes complexas / análise de vínculos;

2.1.2.2.3.32. Cálculo de coeficiente de agrupamento de nós e redes completas;

2.1.2.2.3.33. Cálculo de métrica de centralidade de nós (HITS, PageRank, grau, closeness, betweenness, etc);

2.1.2.2.3.34. Cálculo de métrica de centralidade de arestas (betweenness);

2.1.2.2.3.35. Detecção de comunidades;

2.1.2.2.3.36. Cálculo de assertividade (grau de homofilia);

2.1.2.2.3.37. Representação de afiliações (ou tipos diferentes de nós);

2.1.2.2.3.38. Detecção de balanceamento em redes polarizadas;

2.1.2.2.3.39. Classificação de nós (LOC - Link-based Object Classification);

2.1.2.2.3.40. Predição de links;

2.1.2.2.3.41. Geração de redes (aleatórias, de mundo pequeno, livres de escala);

2.1.2.2.3.42. Cálculo de menor caminho (Dijkstra, A\*, Bellman-Ford, Johnson, PCV, dentre outros);

2.1.2.2.3.43. Cálculo de fluxo de custo mínimo (Ford-Fukerson, Edmonds-Karp, dentre outros);

2.1.2.2.3.44. Cálculo de árvore geradora mínima (Kruskal, Prim, Borůvka, dentre outros);

2.1.2.2.3.45. Busca em grafo (largura, profundidade);

2.1.2.2.3.46. Isomorfismo de grafo;

2.1.2.2.3.47. Ordenação topológica.

#### 2.1.2.2.4. Integração

2.1.2.2.4.1. Possibilitar a importação de dados no seguinte formato: Volumes locais SSD/MVNe;

2.1.2.2.4.2. Possibilitar a importação de dados no seguinte formato: Banco relacional via ODBC;

2.1.2.2.4.3. Possibilitar a importação de dados no seguinte formato: Banco relacional via JDBC;

2.1.2.2.4.4. Possibilitar a importação de dados no seguinte formato: Elasticsearch;

2.1.2.2.4.5. Possibilitar a importação de dados no seguinte formato: MongoDB;

2.1.2.2.4.6. Possibilitar a importação de dados no seguinte formato: Cassandra;

2.1.2.2.4.7. Possibilitar a importação de dados no seguinte formato: XML;

2.1.2.2.4.8. Possibilitar a importação de dados no seguinte formato: JSON;

2.1.2.2.4.9. Possibilitar a importação de dados no seguinte formato: TXT;

2.1.2.2.4.10. Possibilitar a importação de dados no seguinte formato: CSV;

2.1.2.2.4.11. Possibilitar a importação de dados no seguinte formato: Planilha Excel;

2.1.2.2.4.12. Possibilitar a importação de dados no seguinte formato: Planilha ODS;

2.1.2.2.4.13. Possibilitar a importação de dados no seguinte formato: Documento Word;

2.1.2.2.4.14. Possibilitar a importação de dados no seguinte formato: Documento ODT;

2.1.2.2.4.15. Possibilitar a importação de dados no seguinte formato: Documento PDF;

2.1.2.2.4.16. Possibilitar a importação de dados pelo seguinte meio: HTTP;

2.1.2.2.4.17. Possibilitar a importação de dados pelo seguinte meio: HTTPS;

2.1.2.2.4.18. Possibilitar o acesso (leitura e escrita) aos dados no seguinte meio: FTP;

2.1.2.2.4.19. Possibilitar o acesso (leitura e escrita) aos dados no seguinte meio: S3;

2.1.2.2.4.20. Possibilitar o acesso (leitura e escrita) aos dados no seguinte meio: CEPH;

2.1.2.2.4.21. Possibilitar a importação de dados pelo seguinte meio: Web Service SOAP/XML;

2.1.2.2.4.22. Possibilitar a importação de dados pelo seguinte meio: Web Service REST/JSON;

2.1.2.2.4.23. Possibilitar a importação de dados pelo seguinte meio: Páginas Web via Web Crawler;

2.1.2.2.4.24. Possibilitar a importação de dados pelo seguinte meio: Twitter;

2.1.2.2.4.25. Possibilitar a importação de dados pelo seguinte meio: Facebook;

2.1.2.2.4.26. Possibilitar a importação de dados pelo seguinte meio: LinkedIn;

2.1.2.2.4.27. Possibilitar a importação de dados pelo seguinte meio: Quora;

2.1.2.2.4.28. Possibilitar a importação de dados pelo seguinte meio: StackOverflow;

2.1.2.2.4.29. Possibilitar o acesso (leitura e escrita) aos dados no seguinte meio: Hadoop - HDFS;

2.1.2.2.4.30. Possibilitar o acesso (leitura e escrita) aos dados no seguinte meio: Hadoop - Hive;

2.1.2.2.4.31. Possibilitar o acesso (leitura e escrita) aos dados no seguinte meio: Hadoop - Hbase;

2.1.2.2.4.32. Possibilitar o acesso (leitura e escrita) aos dados no seguinte meio: Hadoop - Impala;

2.1.2.2.4.33. Permitir execução e submissão de jobs Spark e MapReduce para serem executados na infraestrutura Hadoop;

2.1.2.2.4.34. Possibilitar o acesso (leitura e escrita) aos dados no seguinte meio: Exadata;

2.1.2.2.4.35. Possibilitar o acesso (leitura e escrita) aos dados no seguinte meio: Teradata;

2.1.2.2.4.36. Possibilitar o acesso (leitura e escrita) aos dados no seguinte meio: Adabas;

2.1.2.2.4.37. Possibilitar o acesso (leitura e escrita) aos dados no seguinte meio: NFS;

- 2.1.2.2.4.38. Possibilidade de publicar um modelo criado para ser consumido na forma de um serviço REST;
- 2.1.2.2.4.39. Possibilidade de publicar um modelo criado no formato Predictive Model Markup Language (PMML);
- 2.1.2.2.4.40. Permitir a extensão da ferramenta (implementação/construção de funções e/ou tarefas);
- 2.1.2.2.4.41. Permitir integração com a ferramenta Spark;
- 2.1.2.2.4.42. Permitir integração dos fluxos de trabalho com a linguagem de programação Python;
- 2.1.2.2.4.43. Permitir integração dos fluxos de trabalho com a linguagem de programação R;
- 2.1.2.2.4.44. Permitir integração dos fluxos de trabalho com a linguagem de programação Scala;
- 2.1.2.2.4.45. Permitir integração dos fluxos de trabalho com a linguagem de programação Java;
- 2.1.2.2.4.46. Permitir integração dos fluxos de trabalho com a linguagem de programação C/C++;
- 2.1.2.2.4.47. Permitir integração dos fluxos de trabalho com a linguagem de programação Julia;
- 2.1.2.2.4.48. Permitir integração dos fluxos de trabalho com a linguagem de programação LUA.
- 2.1.2.2.5. Gestão de configuração
  - 2.1.2.2.5.1. Realizar versionamento dos dados e modelos;
  - 2.1.2.2.5.2. Permitir o versionamento de dados, modelos e visualizações no Git.
- 2.1.2.2.6. Operação
  - 2.1.2.2.6.1. Permitir a distribuição dos modelos criados nos seguintes ambientes (testes, validação, homologação e produção);
  - 2.1.2.2.6.2. Permitir a monitoração do desempenho dos modelos em uso nos ambientes produtivos.
- 2.1.2.2.7. Autenticação e controle de acesso
  - 2.1.2.2.7.1. Permitir a autenticação e autorização via LDAP (Lightweight Directory Access Protocol);
  - 2.1.2.2.7.2. Permitir a autenticação e autorização via Kerberos;
  - 2.1.2.2.7.3. Permitir a autenticação e autorização via OpenID;
  - 2.1.2.2.7.4. Permitir a autenticação e autorização via SAML (Security Assertion Markup Language);
  - 2.1.2.2.7.5. Permitir a autenticação e autorização via Active Directory;
  - 2.1.2.2.7.6. Possibilitar a autenticação Single-Sign-On (SSO);
  - 2.1.2.2.7.7. Permitir a autenticação com usuário e senha via Certificação Digital.
- 2.1.2.2.8. Logs
  - 2.1.2.2.8.1. Possuir logs de sistema;
  - 2.1.2.2.8.2. Possuir logs de acesso por usuário;
  - 2.1.2.2.8.3. Possuir logs de uso das funcionalidades por usuário;
  - 2.1.2.2.8.4. Permitir auditoria dos logs;
  - 2.1.2.2.8.5. Permitir a exportação dos logs.
- 2.1.2.2.9. Usabilidade
  - 2.1.2.2.9.1. Possibilitar a aplicação de técnicas e métodos de processamento e análise de dados e inteligência artificial (através de workflow, componentes, configurações de operações e parâmetros) pelos próprios usuários da ferramenta por meio de recursos visuais e operações de "drag and drop";
  - 2.1.2.2.9.2. Possibilitar a criação dos fluxos de processamento e uso de componentes de forma visual (fontes de dados, preparação dos dados, execução do modelo, avaliação do modelo, publicação).
- 2.1.2.2.10. Visualização e Exportação de Dados
  - 2.1.2.2.10.1. Possuir funcionalidades para visualização dos dados/resultados no seguinte tipo: texto;
  - 2.1.2.2.10.2. Possuir funcionalidades para visualização dos dados/resultados no seguinte tipo: imagem;

- 2.1.2.2.10.3. Possuir funcionalidades para visualização dos dados/resultados no seguinte tipo: tabela;
- 2.1.2.2.10.4. Possuir funcionalidades para visualização dos dados/resultados no seguinte tipo: tabela dinâmica;
- 2.1.2.2.10.5. Possuir funcionalidades para visualização dos dados/resultados no seguinte tipo: métrica / KPI;
- 2.1.2.2.10.6. Possuir funcionalidades para visualização dos dados/resultados no seguinte tipo: histograma;
- 2.1.2.2.10.7. Possuir funcionalidades para visualização dos dados/resultados no seguinte tipo: plotagem da caixa;
- 2.1.2.2.10.8. Possuir funcionalidades para visualização dos dados/resultados no seguinte tipo: plotagem de distribuição;
- 2.1.2.2.10.9. Possuir funcionalidades para visualização dos dados/resultados no seguinte tipo: mapa geográfico;
- 2.1.2.2.10.10. Possuir funcionalidades para visualização dos dados/resultados no seguinte tipo: gráfico de barra vertical;
- 2.1.2.2.10.11. Possuir funcionalidades para visualização dos dados/resultados no seguinte tipo: gráfico de barra horizontal;
- 2.1.2.2.10.12. Possuir funcionalidades para visualização dos dados/resultados no seguinte tipo: gráfico de linha;
- 2.1.2.2.10.13. Possuir funcionalidades para visualização dos dados/resultados no seguinte tipo: gráfico de dispersão;
- 2.1.2.2.10.14. Possuir funcionalidades para visualização dos dados/resultados no seguinte tipo: gráfico de pizza;
- 2.1.2.2.10.15. Possuir funcionalidades para visualização dos dados/resultados no seguinte tipo: gráfico em cascata;
- 2.1.2.2.10.16. Possuir funcionalidades para visualização dos dados/resultados no seguinte tipo: mapa de calor;
- 2.1.2.2.10.17. Possuir funcionalidades para visualização dos dados/resultados no seguinte tipo: mapa de árvore (treemap);
- 2.1.2.2.10.18. Possuir funcionalidades para visualização dos dados/resultados no seguinte tipo: gauge (velocímetro);
- 2.1.2.2.10.19. Possuir funcionalidades para visualização dos dados/resultados no seguinte tipo: tag cloud;
- 2.1.2.2.10.20. Possuir funcionalidades para visualização dos dados/resultados no seguinte tipo: confusion matrix;
- 2.1.2.2.10.21. Possuir funcionalidades para visualização dos dados/resultados no seguinte tipo: plotagem de caixa condicional;
- 2.1.2.2.10.22. Possuir funcionalidades para visualização dos dados/resultados no seguinte tipo: matriz de dispersão;
- 2.1.2.2.10.23. Possuir funcionalidades para visualização dos dados/resultados no seguinte tipo: plotagem de radar;
- 2.1.2.2.10.24. Possuir funcionalidades para visualização dos dados/resultados no seguinte tipo: gráfico misto (combinação de outros tipos);
- 2.1.2.2.10.25. Possibilitar a exportação de visualização dos dados/resultados no formato PDF;
- 2.1.2.2.10.26. Possibilitar a exportação de visualização dos dados/resultados no formato SVG;
- 2.1.2.2.10.27. Possibilitar a exportação de visualização dos dados/resultados no formato GIF;
- 2.1.2.2.10.28. Possibilitar a exportação de visualização dos dados/resultados no formato PNG;
- 2.1.2.2.10.29. Possibilitar a exportação de visualização dos dados/resultados no formato JPG;
- 2.1.2.2.10.30. Possibilitar a exportação de visualização dos dados/resultados no formato EPS;
- 2.1.2.2.10.31. Possibilitar a exportação de visualização dos dados/resultados no formato Excel;
- 2.1.2.2.10.32. Possibilitar a exportação de visualização dos dados/resultados no formato ODS;
- 2.1.2.2.10.33. Possibilitar a exportação de visualização dos dados/resultados no formato HTML;
- 2.1.2.2.10.34. Possibilitar a exportação de visualização dos dados/resultados no formato HTML5 e JavaScript;
- 2.1.2.2.10.35. Possibilitar a exportação de visualização dos dados/resultados no formato Word;
- 2.1.2.2.10.36. Possibilitar a exportação de visualização dos dados/resultados no formato ODT;
- 2.1.2.2.10.37. Possibilitar a exportação de visualização dos dados/resultados no formato CSV;
- 2.1.2.2.10.38. Possibilitar a exportação de visualização dos dados/resultados no formato XML.

#### 2.1.2.2.11. Ambiente

2.1.2.2.11.1. Permitir execução, pelo menos, nas plataformas Linux e Windows;

2.1.2.2.11.2. Suporte à execução concorrente de fluxos de trabalho;

2.1.2.2.11.3. Suporte ao treinamento dos modelos em paralelo (multithread) em CPU;

2.1.2.2.11.4. Suporte à execução dos modelos (inferência) em paralelo (multithread) em CPU;

2.1.2.2.11.5. Suporte ao treinamento dos modelos de forma distribuída (em mais de um servidor) em CPU;

2.1.2.2.11.6. Suporte à execução dos modelos (inferência) de forma distribuída (em mais de um servidor) em CPU;

2.1.2.2.11.7. Suporte ao treinamento dos modelos em paralelo (multithread) em GPU CUDA;

2.1.2.2.11.8. Suporte à execução dos modelos (inferência) em paralelo (multithread) em GPU CUDA;

2.1.2.2.11.9. Suporte ao treinamento dos modelos de forma distribuída (em mais de um servidor) em GPU CUDA;

2.1.2.2.11.10. Suporte à execução dos modelos (inferência) de forma distribuída (em mais de um servidor) em GPU CUDA;

2.1.2.2.11.11. Suporte ao treinamento e execução dos modelos de forma distribuída em clusters Hadoop;

2.1.2.2.11.12. Suporte à arquitetura x86 64bits;

2.1.2.2.11.13. Suporte à arquitetura POWER;

2.1.2.2.11.14. Suporte a containers Docker;

#### 2.1.3. Requisitos de Implantação

2.1.3.1. O produto deverá ser implantado 100% na infraestrutura do SERPRO, com administração pelos técnicos do SERPRO, não devendo trafegar nenhum dado para nuvens externas.

2.1.3.2. Todas a infraestrutura de hardware e software básico (sistemas operacional e gerenciador de banco de dados) serão providos pelo SERPRO.

#### 2.1.4. Modelos de Comercialização

2.1.4.1. O produto deverá ser ofertado com modelos de comercialização na modalidade de pagamento pelo uso e na modalidade de licenciamento de uso perpétuo.

### 3.0 Níveis de serviço e sancionamentos

3.1. As empresas devem fornecer documento com a descrição dos níveis de serviços inerentes ao produto ofertado.

### 4.0 Especificação de valores e forma de pagamento

4.1. As empresas devem fornecer documento com a descrição dos modelos de comercialização na modalidade de pagamento pelo uso e na modalidade de licenciamento de uso perpétuo.

### 5.0 Justificativa da contratação

5.1. Esta Consulta Pública está autorizada pelo Diretor da DIOPE por meio do SISCOR SUPES 020760/2019-25 (cópia em anexo).

5.2. A Consulta Pública tem com objetivo validar junto ao mercado os requisitos necessários para contratação de um produto para construção de soluções de inteligência artificial por meio de automated machine learning (AutoML), que possibilitem ao SERPRO construir soluções de IA com alta produtividade e em larga escala, além disso, ofertar para seus clientes uma plataforma tecnológica que possibilite aos seus especialistas a criação de suas próprias soluções.

5.3. O Produto para construção de soluções de inteligência artificial visa atender à demanda já formalizada da Receita Federal do Brasil - RFB - e possibilitar a criação de produtos SERPRO a serem ofertados para o mercado em geral.

### 6.0 Seleção do fornecedor

6.1. Consulta pública com fulcro no Art. 31, da Lei nº 9.784/1999, objetivando esclarecimentos sobre produtos, processos, soluções e tecnologias junto ao mercado.

### 7.0 Justificativa para aceitação de preços



7.1. Não de aplica.

## 8.0 Gerenciamento contratual

8.1. A Consulta Pública Eletrônica será acompanhada pelos empregados:

8.1.1. Gustavo Monti Rocha, telefone (31)3311-6214, e-mail: gustavo.rocha@serpro.gov.br

8.1.2. Carlos Rodrigo Fonseca Lima, telefone (31)3311-6943, e-mail: carlos-rodrigo.lima@serpro.gov.br

8.1.3. Jones Lamanna Tesser, telefone (11)3730-1034, e-mail: jones.tesser@serpro.gov.br

8.1.4. Leandro Dutra e Silva, telefone (61) 2021-8726, e-mail: leandro.silva@serpro.gov.br

## 9.0 Considerações gerais

9.1. Não de aplica.

## Elaboração

Data : 30/07/2019

LEANDRO DUTRA E SILVA - 21072167

SUPES/ESTAD/ESECT

Anexos

Arquivo: [Siscor de Autorização](#)

Arquivo: [Siscor Autorizado](#)

..