Fatec
São José dos
Campos
Prof. Jessen Vidal

PROJETO KERSYS: Aplicação de Técnicas de Machine Learning para Auxiliar na Tomada de Decisão e Avaliação do Negócio Florestal.

Aline Oliveira, Caroline Nunes, Débora Santos e Larissa Janine

Análise e Desenvolvimento de Sistemas, 6º Período
Profº José Walmir Duque | Disciplina: Inteligência Artificial,
São José dos Campos, São Paulo, Brasil, 02 de outubro de 2020

SUMÁRIO

INTRODUÇÃO

As florestas são uma fonte de recursos naturais e desempenham um importante papel na preservação de um ambiente sustentável para a vida humana. O reconhecimento da importância multifuncional destes ecossistemas, como produtores de madeira, resina, biomassa, e sobretudo de bens indiretos como o turismo, a reposição da biodiversidade, regulação dos fluxos de água, controle de erosão, sumidouro de carbono, ou então simplesmente pela produção de oxigénio, colocam novos desafios aos profissionais que trabalham ou gerem estes espaços. Também ao nível político tem havido uma maior consciencialização do seu papel, levando a que, por iniciativa própria ou por acordos internacionais, tenha surgido um conjunto de legislação e de instrumentos de planeamento e ordenamento que estipulam, regulam e condicionam a sua atividade de forma espacial, tendo em vista a preservação e sustentabilidade dos ecossistemas [1].

É nesta matéria que o planejamento para gestão florestal se faz extrema importância para a verificação desde simples produções de cartográficas temáticas, com delimitação e enquadramento das diferentes propriedades florestais, até estudos mais complexos, como a elaboração de cartografias de aptidão de espécies florestais ou cartografia de risco de incêndio.

O uso das tecnologias para suas capacidades de armazenamento, integração, edição, extração, visualização e análise de diferentes tipos de dados georreferenciados, permitem um conhecimento mais concreto e preciso das situações, criando informação atualizada e facilitando a tomada de decisão dos diversos intervenientes, desde governos, passando pelos gestores, técnicos e pelos próprios proprietários florestais [1].

E atuante neste mercado, a Kersys é uma empresa de tecnologia da informação especializada em prover soluções para otimizar a Gestão Florestal e do Agronegócio, desenvolvendo sistemas informatizados para planejamento, controle e gestão de máquinas e plantações, promovendo a capacitação dos clientes para a operação e utilização eficiente dos seus recursos.

O presente projeto tem como principal meta, a utilização de um algoritmo de machine learning para calcular a projeção de produtividade a partir de padrões identificados entre as variáveis do banco de dados, qual foi-se construído baseado na inserção de informações de atividades do campo.

1. BUSINESS UNDERSTANDING

O projeto KIA

A empresa possui um projeto em desenvolvimento denominado KIA, cujo objetivo é desenvolver análises inteligentes a partir de um banco de dados com informações operacionais do campo, de clima, geográficas e de manejo, visando a maximização da produtividade florestal.

The Challenge

O projeto para a empresa Kersys trata-se da elaboração de um algoritmo para calcular inteligentemente a projeção de produtividade a partir do padrão identificado entre as variáveis da base de dados. A realização destas análises de projeção de resultados de produtividade será realizada com o auxílio da Inteligência Artificial, visando a assertividade nas recomendações para futuras tomadas de decisão.

Desta forma, será possível que usuários do sistema insiram informações operacionais ao longo do tempo e possuam expectativas de resultados de produtividade a serem atingidos a partir de suas ações no campo, antes da obtenção da produtividade real da floresta plantada.

ATORES

Empresa	Participante	Função
KERSYS	Ana Melo	-
KERSYS	José Roberto	-
FATEC	Débora Santos	Master
FATEC	Caroline Nunes	PO
FATEC	Aline de Oliveira	DEV
FATEC	Larissa Janine	DEV

Tarefas elegíveis a priori

Todos os dados, processos e métricas necessários precisam estar disponíveis no formato requerido.

- Áreas colocaremos em metros, porcentagens em porcentagem, texto em texto, etc. Pois sem essa conversão a visualização e interpretação dos dados por nós e pela Kersys ficará mais complexa e as predições não assertivas.
- Precisamos também tratar os dados, procurar por *null* values e fazer o preenchimento ou a eliminação de tais células.
- Armazenar o documento .xml como csv e .arff

Requisitos Funcionais

- Considerando que o banco de dados relacional já pronto, espera-se que sejam elaborados algoritmos de identificação dos padrões que ocorrem entre as variáveis descritivas e de resultado do banco de dados.
- Espera-se a elaboração de algoritmo para calcular a projeção de produtividade a partir do padrão identificado entre as variáveis do banco de dados, com base na inserção de informações de atividades do campo.
- Será feita o acompanhamento dos resultados a fim de validá-los com o conhecimento agroflorestal e identificar se os resultados estão de acordo com o que poderia ser observado no campo, o que poderá permitir a criação de mecanismos de feedback do próprio algoritmo.
- Será passado estudo pré-estabelecido das features a serem consideradas (banco de dados já está estruturado).

Requisitos Não Funcionais

- Linguagem Python
- Biblioteca SKLearn e/ou outras aplicáveis a cada Projeto
- Banco de Dados Relacional ou NoSQL
- SGBD Microsoft SQL Server

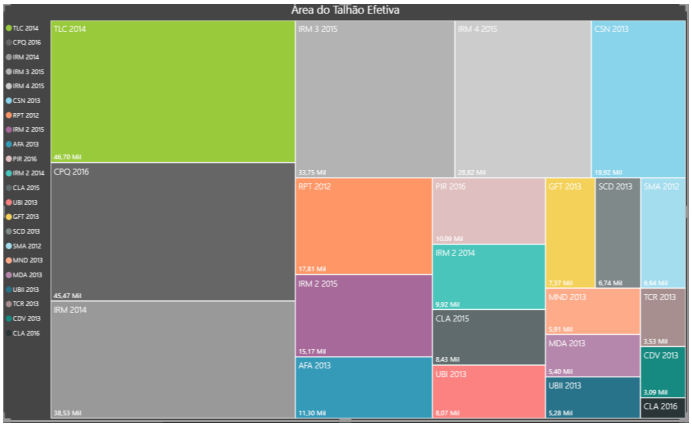
2. DATA UNDERSTANDING

Com o banco de dados relacional já pronto, constituído por dados obtidos pela própria empresa Kersys no decorrer dos anos, analisando e acompanhando as áreas que administram, exploraremos os dados, filtrando-os apenas os que forem relevantes para desenvolvimento do projeto e verificando a qualidade dos mesmos.

A compreensão dos dados já existentes é a principal ação para extrair informações consistentes para a elaboração do algoritmo que calculará a projeção de produtividade futura. Levando em conta que novos dados serão inseridos por usuários ao longo do tempo, um processo de seleção dos dados já estará estabelecido e novas projeções serão calculadas.

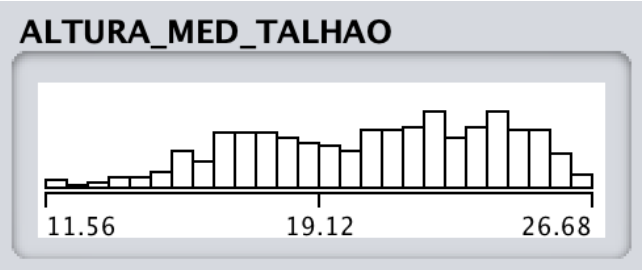
Dentre os dados que são relevantes para tal funcionalidade, destacam-se:

- Área plantada (533 talhões) variando entre 0,44 e 85.29 ha. (**AREA_TALHAO**)



- Produtividade obtida em cada safra; (**FUSTES_HA**)
- Tempo de cada ciclo; (**CICLO**)
- Disponibilidade de água no solo (Categorias: até 20%, 20 - 40% | 40 - 60% | 60 – 80% | 80 – 100%)
- Condições climáticas do local; (**CHOVEU?**) se sim, (**QUANTO**?) (*Categorias: 1 - Chuva até 1 mm; 2 - Entre 1 e 5 mm; 3 - Entre 5 e 15 mm; 4 - Entre 15 e 30 mm; 5 - Acima de 30 mm.)*
- Ano da Atividade Realizada: Entre 2012 e 2018
- Ano do Plantio: Entre 2012 e 2016;
- Idade do Plantio;
- Acréscimo Anual;
- Quantidade de Fustes x Tamanho do Hectare;

- Perdas consideráveis; (**MORTES_PERC**);
- Tipo de Atividades (Manual ou Mecânica) (**DESC_GRP_ATIVIDADE**)
- Materiais Genéticos: 15 tipos distintos;
- Altura Média do talhão (em metros):



- Entre outros.

E dentre todas as variáveis existentes, utilizaremos das seguintes para resolução do problema:

Variável	Tipo da Variável	Descrição da Variável
AREA_TALHAO	Descritiva	Área do talhão, em hectares.
CICLO	Descritiva	Ciclo de produção em que se encontra o talhão. Entende-se como ciclo o período produtivo de um talhão após o plantio de mudas. Um ciclo pode ter mais de uma rotação. No caso deste banco de dados, todos os talhões estão em primeiro ciclo. Após a colheita e com novo plantio, o talhão entrará em segundo ciclo de produção.
ROTACAO	Descritiva	Rotação corresponde ao manejo da área após colheita, mas sem novo plantio de mudas, com crescimento e condução da rebrota dos tocos que permaneceram na área.
SIGLA_MAT_GEN	Descritiva	Descrição do material genético cultivado no talhão, ou seja, identifica a variedade e espécie utilizada.
DATA_PLANTIO	Descritiva	Data em que as mudas foram plantadas no talhão.
MES_PLANTIO	Descritiva	Mês em que as mudas foram plantadas no talhão.
ANO_PLANTIO	Descritiva	Ano em que as mudas foram plantadas no talhão.

DESC_GRP_ATIVIDADE	Descritiva	Descreve o grupo de atividades da operação realizada no campo. Por exemplo, grupo "Fertilização" corresponde às práticas de adubação, e incluem atividades como "Adubação Mecânica Cobertura - 12 meses" e "Adubação Mecânica Cobertura - 6 meses", realizadas em períodos e de formas diferentes
NOME_ATIVIDADE	Descritiva	Descreve a operação efetivamente realizada no campo.
MODO_DE_OPERACAO	Descritiva	Descreve o modo de operação da atividade realizada, se feita de forma mecânica, manual, aérea ou se corresponde à irrigação.
MODO_DE_APLICACAO	Descritiva	Descreve o modo de aplicação de eventuais insumos utilizados na atividade, se em Área Total (toda a extensão do talhão), Localizada (próximo às linhas de plantio, apenas), Cova (apenas na cova de plantio da muda) ou Irrigação.
MODO_DE_ACAO	Descritiva	Descreve o modo de ação da atividade, se ela atinge seus objetivos por vias químicas, mecânicas ou irrigação.
QTDE_REALIZADA_ATIVIDADE	Descritiva	Especifica a área do talhão em que a atividade foi efetivamente realizada. Pode ser diferente da área total do talhão.
IDADE_TALHAO_ATIVIDADE	Descritiva	Informa a idade do plantio no momento de realização da atividade no campo.
VOLUME_HA_TALHAO	Resposta	Volume de madeira por hectare no talhão - m^3/ha
FUSTES_HA	Resposta	Corresponde ao número de fustes por hectare no talhão, ou seja, número de árvores viáveis para colheita.

MORTES_PERC	Resposta	Corresponde ao percentual de mortes de fustes no talhão (%), ou seja, árvores não produtivas.
CHOVEU?	Descritiva	Variável booleana que aponta se choveu no dia de execução da atividade no campo (DATA_ATIVIDADE_REALIZ)
QUANTO?	Descritiva	Variável categórica que aponta o volume de chuvas no dia de execução da atividade (DATA_ATIVIDADE_REALIZ). Categorias: 1 - Chuva até 1 mm; 2 - Entre 1 e 5 mm; 3 - Entre 5 e 15 mm; 4 - Entre 15 e 30 mm; 5 - Acima de 30 mm.
C.DIA ANT?	Descritiva	Variável booleana que aponta se choveu no dia anterior à execução da atividade no campo (DATA_ATIVIDADE_REALIZ)
C.DIA POST?	Descritiva	Variável booleana que aponta se choveu no dia posterior à execução da atividade no campo (DATA_ATIVIDADE_REALIZ)
C.3 DIAS A?	Descritiva	Variável booleana que aponta se choveu nos três dias anteriores à execução da atividade no campo (DATA_ATIVIDADE_REALIZ)
C.3 DIAS D?	Descritiva	Variável booleana que aponta se choveu nos três dias posteriores à execução da atividade no campo (DATA_ATIVIDADE_REALIZ)

Variável Objetivo e Classes

A Variável Objetivo será a FUSTES_HA, onde indicará se o número de árvores viáveis para colheita.

As Classes serão “PRODUÇÃO_EXCELENTE”, na qual representará se a quantidade de fustes colhidos por talhão foi mais de 90% do plantio, “PRODUÇÃO_MEDIANA”, na qual representará se a quantidade de fustes colhidos por talhão foi mais de 50% do plantio, e “PRODUÇÃO_RUIM”, na qual representará se a quantidade de fustes colhidos por talhão foi menos de 50% do plantio.

Possíveis Contratempos com os Dados

- Acontecimentos incontroláveis pela empresa, como um desastre natural, queimadas, acidentes, entre outros. Portanto, destaca-se a importância de notificar quando esses acontecimentos afetarem o resultado, para que seja considerado dentre os algoritmos.
- Quantidade média de informação e por quanto tempo o cliente deverá imputar informações para que haja uma predição.
- De quanto tempo será a predição? 6 meses, 1 ano, 5 anos, 20 anos?
- Erros na execução das atividades, podendo ser errada a data, o modo de operação, à aplicação de insumos, entre outros.
- Inserção de dados incorretos pelo usuário no banco de dados

*Problemas como os listamos acima refletem a importância da etapa a seguir.

3. DATA PREPARATION

Tarefas de Mineração de dados

Tarefa 1: Classificação

Justificativa de escolha: Essa tarefa nos auxiliará a construir um modelo de algum tipo específico de dado que possa ser aplicado a dados não classificados a fim de categorizá-los em classes, o objetivo é descobrir um relacionamento entre um atributo meta (cujo valor será previsto) e um conjunto de atributos de previsão. Usaremos essa tarefa para classificar por exemplo:

Meta: Influência da fertilização na altura média do talhão.

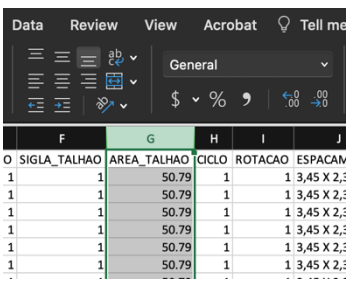
Atributos de previsão: Fertilização, Altura Média do Talhão

Tarefa 2: Estimativa

Justificativa de escolha: Esta tarefa será utilizada para definirmos um valor para alguma variável contínua desconhecida, podemos estimar o valor de uma determinada variável analisando-se os valores das demais. É uma das tarefas mais utilizadas e importantes para o processo de mineração de dados, no nosso projeto, pode ser usado para estimar a probabilidade de que uma árvore morrerá nos baseando nos resultados de tipos de ações tomadas e quantidade de chuva, por exemplo.

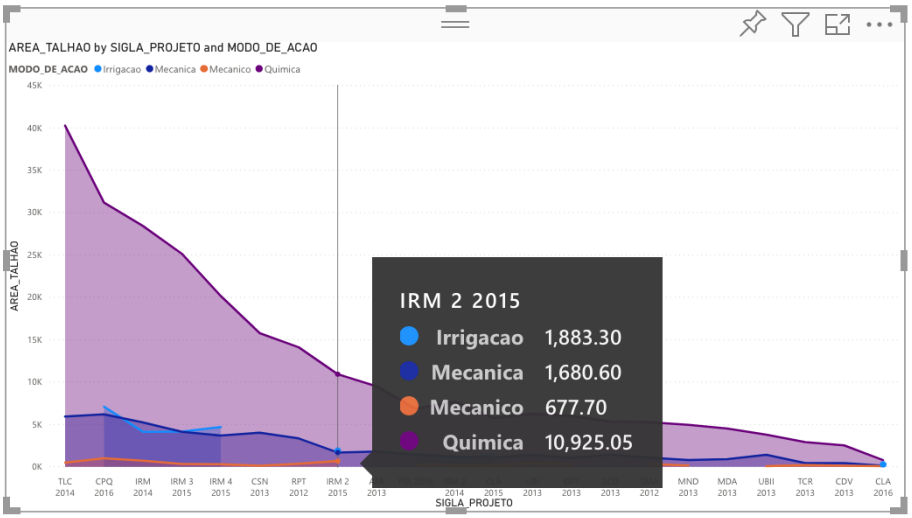
Para a etapa de preparação dos dados, nosso banco de dados reestruturado encontra-se no repositório do nosso Github. A junção dos dados já foi feita, pois o banco de dados veio concentrado em uma planilha só. A avaliação dos dados está continuamente em processo, já a limpeza dos dados está sendo feita conforme vamos correlacionando as variáveis para obtermos um determinado resultado, quando não é possível, verificamos a fonte e a descrição dos dados.

A transformação e enriquecimento dos dados têm sido de extrema importância, pois quando analisamos, por exemplo Área talhão X Fustes. Fustes está representado por quantidade e área por hectares, então se não adequamos a variável **fustes_ha** a análise em questão, obtemos valores exorbitantes, fora da realidade. Note que o tipo de dado consta como Geral, portanto, para obtermos uma boa comparação, é necessário a utilização de fórmulas, para cálculo da porcentagem.



Note a descrição dos valores dentro da variável **MODO_DE_AÇÃO:** "*Descreve o modo de ação da atividade, se ela atinge seus objetivos por vias químicas, mecânicas ou irrigação."* Quando comparamos a **AREA_TALHAO** x **MODO_DE_ACAO**, são apresentadas QUATRO categorias, sendo que Mecânica e Mecânico são similares, acreditamos que todos os valores de nome "Mecânico" deverão assumir o nome de "Mecânica".

Outro problema é que segundo a análise do plot abaixo, os talhões de área menor não usam a **irrigação** como um **modo de ação**?



Considerações Finais

Problemas como estes e outros são imprescindíveis para a transformação e enriquecimento dos dados, tanto que, apenas por mudar o tipo do dado e como ele é representado, a interpretação da análise fica imediatamente mais intuitiva e assertiva, satisfazendo tanto os experts quanto os leigos no ramo de produção florestal e é exatamente isso que queremos atingir ao final deste projeto, facilitar a interpretação dos dados por todos e qualquer usuário do sistema da Kersys.

GLOSSÁRIO

- **Algoritmo:** É uma sequência ordenada, definida e finita de ações que visam a solução de um determinado problema computacional. Em suma, o problema contém um conjunto de dados de entrada (input) e o algoritmo, na sequência das ações resolventes, produz os dados de saída (output).
- **Banco de Dados Relacional:** A linguagem padrão dos Bancos de Dados Relacionais é a Structured Query Language, ou simplesmente SQL, como é mais conhecida. Os dados de um banco de dados relacional são armazenados em tabelas. Uma tabela é uma simples estrutura de linhas e colunas. Em uma tabela, cada linha contém um mesmo conjunto de colunas.
- **Biblioteca SKLearn :** Scikit-learn ou Sklearn é uma biblioteca baseada em Python para construir modelos de aprendizado de máquina. Ele fornece muitos algoritmos de regressão, clustering e classificação. O Sklearn é compatível com o NumPy e o SciPy. Isso significa que você poderá interoperar com diferentes bibliotecas Python facilmente.
- **Biomassa:** É um recurso renovável proveniente de matéria orgânica, que tem por objetivo principal a produção de energia. No Brasil a biomassa florestal sempre teve papel importante na matriz energética, tendo como principal uso o carvão vegetal e a lenha.
- **Business Understanding:** Foca em entender o objetivo do projeto a partir de uma perspectiva de negócios, definindo um plano preliminar para atingir os objetivos.
- **Ecossistemas:** O ecossistema é definido como sendo o conjunto formado por comunidades bióticas que habitam e interagem em determinada região e pelos fatores abióticos que exercem influência sobre essas comunidades.
- **Requisitos Funcionais:** Está se referindo à requisição de uma função que o software terá que atender/realizar. Ou seja, exigência, solicitação, desejo, necessidade, em que o software deverá materializar.
- **Requisitos Não Funcionais:** São os requisitos relacionados ao uso da aplicação em termos de desempenho, usabilidade, confiabilidade, segurança, disponibilidade, manutenção e tecnologias envolvidas.

REFERÊNCIAS

GOMES, P.M.M., 2012. Desenvolvimento de um Sistema de Informação e Apoio à Gestão Florestal baseado em Tecnologia Open Source. Dissertação de Mestrado em Sistemas de Informação Geográfica, Universidade de Trás-os-Montes e Alto Douro, Vila Real, 120 pp

Vazquez, Carlos; Simões, Guilherme (2016). Engenharia de Requisitos: Software Orientado ao Negócio. [S.l.]: Brasport

Silva, Afonso da Silva (2013). Direito Ambiental Constitucional. 10. São Paulo: Malheiros. 97 páginas.

Disponível em 02/10/2020: https://www.matanativa.com.br/blog/biomassa-florestal-para-a-geracao-de-energia/#:~:text=Biomassa%20%C3%A9%20um%20recurso%20renov%C3%A1vel,carv%C3%A3o%20vegetal%20e%20a%20lenha

Disponível em 02/10/2020: https://www.infopedia.pt/$algoritmo-(informatica)#:~:text=Um%20algoritmo%20%C3%A9%20uma%20sequ%C3%AAncia,dados%20de%20sa%C3%ADda%20(output).

Disponível em 02/10/2020: https://pt.wikipedia.org/wiki/Cross_Industry_Standard_Process_for_Data_Mining

Disponível em 02/10/2020: https://www.ateomomento.com.br/o-que-e-requisito-funcional/#:~:text=Quando%20falamos%20de%20um%20Requisito,que%20um%20software%20dever%C3%A1%20materializar.

Disponível em 02/10/2020: https://br.bitdegree.org/tutoriais/bibliotecas-python/#:~:text=Scikit%2Dlearn%20ou%20Sklearn%20%C3%A9,com%20diferentes%20bibliotecas%20Python%20facilmente.

Disponível em 02/10/2020: http://dominatudo.blogspot.com/2009/01/conceito-de-banco-de-dados-relacional.html#:~:text=A%20linguagem%20padr%C3%A3o%20dos%20Bancos,SQL%2C%20como%20%C3%A9%20mais%20conhecida.&text=Os%20dados%20de%20um%20banco,um%20mesmo%20conjunto%20de%20colunas.