

# CIÊNCIA DA COMPUTAÇÃO



## Realização de Previsões de Roubo e Furto na Região de São Paulo Utilizando Inteligência Artificial e Estatística

**Alunos:** Anna Beatriz Pereira Lima, Giovana Ellero Viera, Vagner Batazoli Pereira Filho, Vinicius Moura Rodrigues

**Orientadora:** Dra. Leila Cristina C Bergamasco | Leila.Cristina@fei.edu.br

### RESUMO

No cenário atual da região de São Paulo, a persistência dos desafios relacionados a roubos e furtos representa uma inquietação constante. Este trabalho propõe o desenvolvimento de um modelo que, por meio da utilização de dados fornecidos pela Secretaria de Segurança Pública do Estado de São Paulo (SSP-SP), visa antecipar as ocorrências futuras de roubos e furtos na cidade, fundamentado no histórico de crimes na área. A principal finalidade foi criar um modelo capaz de identificar as áreas de maior risco. Para atingir esse objetivo, foram coletados dados de boletins de ocorrência, abrangendo informações como locais, horários, modus operandi e tipologia do crime, os quais foram analisados por meio de técnicas de estatística de Monte Carlo aliadas ao método de inteligência artificial, Random Forest. A expectativa é que este projeto contribua para a identificação da incidência de roubos e furtos na região de São Paulo, fornecendo informações que atribuam valor para a implementação de medidas preventivas pela prefeitura.

### OBJETIVO

Os objetivos deste estudo incluiu o desenvolvimento de um sistema que utiliza dados da SSP-SP para prever futuros crimes, com base no histórico de ocorrências na região de São Paulo. Especificamente, o enfoque principal era a identificação das áreas de maior risco para roubos e furtos, com o intuito de possibilitar a implementação de ações preventivas mais eficazes, visando à melhoria da segurança na região. Adicionalmente, o trabalho buscou atingir os seguintes objetivos específicos:

- Desenvolver um método de previsão de roubos e furtos com base em dados da SSP-SP.
- Identificar as áreas de maior risco para a ocorrência desses crimes.

### METODOLOGIA

A proposta metodológica deste trabalho consistiu em quatro etapas principais. A primeira etapa consistiu no acesso e download da base de dados de criminalidades disponível pelo site da SSP-SP. A segunda etapa foi a de tratamento da base de criminalidade, durante essa fase, os dados brutos foram submetidos a um processo de limpeza, organização e um pré-processamento para análise. Na terceira etapa, foram aplicados os modelos estatísticos (Monte Carlo) e de aprendizado de máquina (Random Forest) para análise de dados e previsões. Essa fase envolveu a configuração, treinamento e avaliação dos modelos, bem como a geração de resultados ou simulações com base nos dados fornecidos. Por fim, foram analisados os resultados obtidos após a aplicação da modelagem. Foram examinadas em detalhes as simulações estatísticas e as previsões de aprendizado de máquina para identificar tendências e padrões relevantes relacionados à criminalidade em São Paulo.

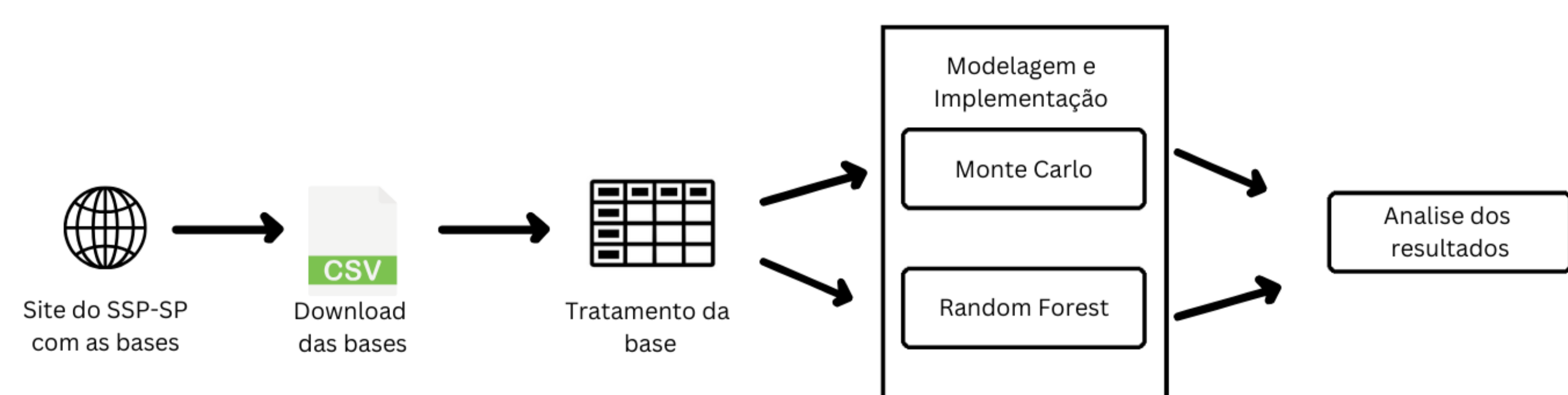


Figura 1 - Fluxograma das etapas da metodologia

### RESULTADOS

Foi aplicado o modelo de Monte Carlo à base de dados de São Paulo e do ABC com os seguintes resultados de desempenho obtidos e também foram gerados heatmaps para visualizar a quantidade de crimes na região do município de São Paulo revelando uma correspondência significativa entre os dois conjuntos de dados.

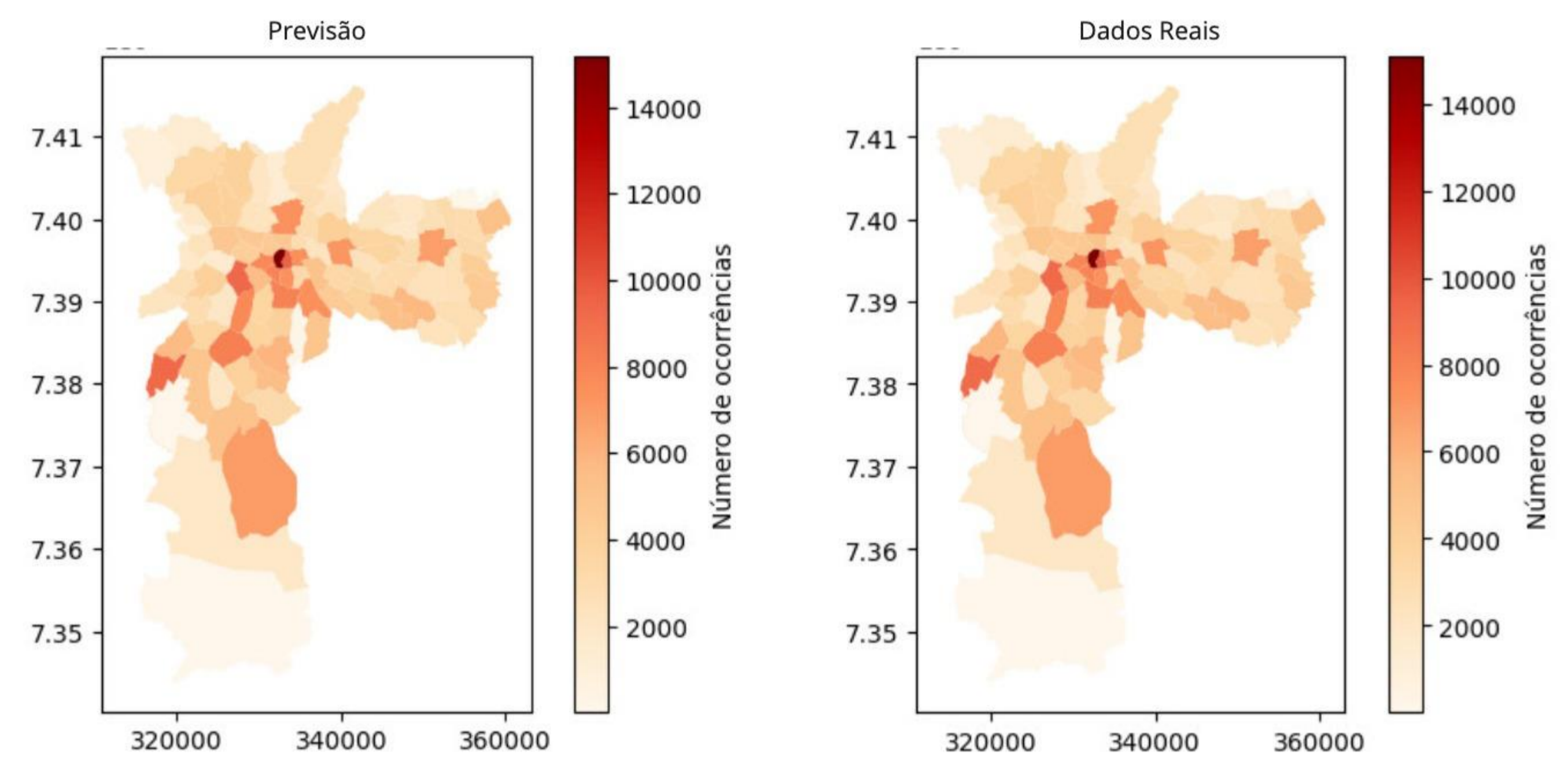


Figura 2 - Heatmap: Quantidade de Crimes no Município de São Paulo

Com a regressão do Monte Carlo, os resultados obtidos da base de dados ABC mostrou uma métrica melhor com a Raiz do Erro Quadrático Médio (RMSE) de 7.0812, superiores em comparação com a base de SP já o Coeficiente de Determinação ( $R^2$ ) da base de SP obteve um desempenho melhor com 0.9999.

Com relação ao modelo Random Forest, foi utilizado suas duas variantes, o Random Forest Regressor e o Random Forest Classifier.

Utilizando o Random Forest Regressor, os resultados obtidos da base de dados ABC mostraram métricas superiores em comparação com a base SP. Com RMSE de 0.3741 e  $R^2$  de 0.4321.

Na tarefa de classificação utilizando o Random Forest Classifier, os resultados obtidos da base de dados SP mostraram métricas superiores em comparação com a base ABC. Especificamente, as métricas para a base SP foram: acurácia de 80%, precisão de 81%, revocação de 77% e F1-Score de 79%. Esses valores indicam uma performance robusta e confiável do modelo para a tarefa de classificação da base SP, superando os resultados da base ABC em todos os aspectos avaliados.

### CONCLUSÃO

Diante dos altos índices de roubos e furtos em São Paulo, este estudo desenvolveu um modelo preditivo com dados da SSP-SP. Utilizando o método estatístico de Monte Carlo e o modelo de IA Random Forest, foi possível realizar previsões consistentes. O Monte Carlo destacou-se na região do ABC, enquanto heatmaps mostraram a correspondência entre crimes previstos e reais. O Random Forest permitiu prever crimes e identificar características geográficas associadas, útil para políticas preventivas. Os resultados indicam um avanço na compreensão da criminalidade em São Paulo, validando a viabilidade de modelos estatísticos e de machine learning na previsão de crimes e na formulação de estratégias de segurança pública e contribuindo com informações que atribuam valor para a implementação de medidas preventivas pela prefeitura.