Machine Learning



João Vitor Pinheiro Nascimento:388837

Emanuel Tosto Holanda Vasconcelos:374172

Centro de Ciências - Departamento de Computação

Índice

1. Introdução............................................................................................3
2. Desenvolvimento..................................................................................4

2.1 Conjunto de dados.........................................................................4

2.2 Análise de dados............................................................................4

1. Conclusão.............................................................................................5
2. Bibliografia...........................................................................................6

1 - Introdução

Vários estudos feitos na economia brasileira mostram uma grande redução da participação da indústria no PIB do país. Estima-se que a desindustrialização do país começou na década de 1980 e atingiu marcos como o peso da indústria de transformação cair de 32,1% do PIB para 19,7% em 1998. Já outros economistas afirmam que as mudanças na economia brasileira nos últimos 30 anos não tiveram nenhum efeito negativo sobre a indústria.

Em meio a tanta discussão surge a motivação desse estudo. Analisar a economia dos estados brasileiros e descobrir quais são focados em qual área da economia por meio dos seus PIBs de indústria, agropecuária e serviços e prever qual os diferentes PIBs com base nas informações dadas.

2 – Desenvolvimento

1. Conjunto de dados:

1.1 Dataset - <https://www.kaggle.com/crisparada/brazilian-cities>

Esse conjunto possui informações sobre população, posses e serviços mais usados de diversos estados brasileiros e como seu PIB é distribuído entre agropecuária, indústria e serviços.

1.2 Tratamento de dados

Para esse estudo foram retirados do dataset informações irrelevantes como nome do estado, nome da cidade e limitar os dados para apenas 2016 (dessa forma mantendo a análise o mais atual possível).

Devido a enorme variância numérica nos dados é necessário normaliza-los para manter a consistência e para que nenhum dado influência de maneira desproporcional os modelos.

1. Análise dos dados:

2.1 Modelo Linear:

Para a análise linear foi utilizado o modelo Gradiente Descendente (GD). Na análise desse modelo ficou claro que certos GVA’s não são linearmente dependentes dos atributos que se espera, como por exemplo o GVA de industrias com relação ao número de companhias de transformação:

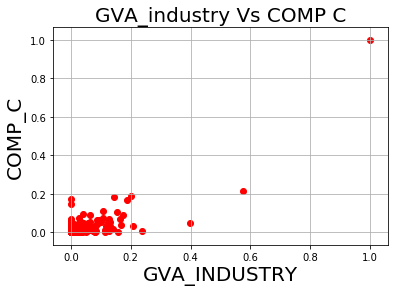


Imagem 1: indústria x companhias de transformação

Em contraste, outros GVA’s funcionam em modelos lineares como o de serviços públicos:

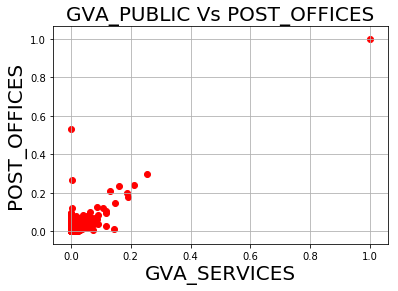


Imagem 2: serviços públicos x agência de correios

Já o GVA de serviços cresce quase que perfeitamente de maneira linear:

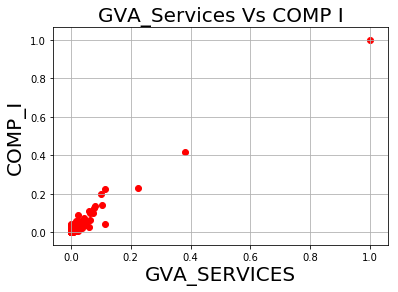


Imagem 3: serviços x companhias de acomodação e comida

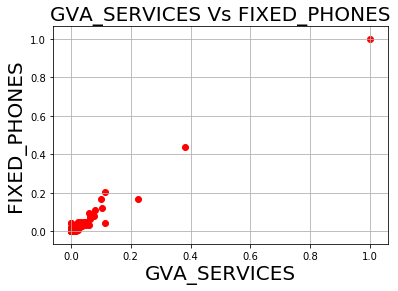
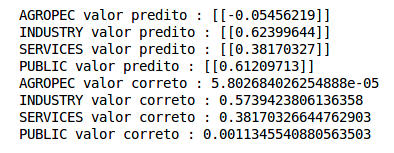


Imagem 4: serviços x telefones fixos

Os erros obtidos refletem bem essa limitação dos modelos lineares. GVA’s de serviços e serviços públicos possuíam erros baixíssimos enquanto GVA’s de industrias e agropecuária possuíam erros muito grandes. Isso pode ser melhor observado em um exemplo de predição:



2.2 Modelo Não Linear:

CONCLUSÃO

Com isso podemos concluir que há uma relação bem direta sobre os tipos de posses, serviços e população e qual o seu PIB mais predominante. Com relação aos modelos os dados em questão são de natureza bem linear e por isso não houve muita vantagem na utilização do MLP.

Essa previsão pode permitir descobrir tendências a certas partes da economia em diferentes locais e por sua vez uma melhor visão de onde deve-se investir mais ou menos.

Bibliografia

## José Luis Oreiro; Carmem A. Feijó (2010) “Desindustrialização: conceituação, causas, efeitos e o caso brasileiro”, artigo da *Brazilian Journal of Political Economy.*

Link -

<http://www.scielo.br/scielo.php?pid=S0101-31572010000200003&script=sci_arttext>