

ANALISANDO MICRODADOS DO ENADE: O Desempenho dos alunos do IC-UFF

Universidade Federal Fluminense

Instituto de Computação

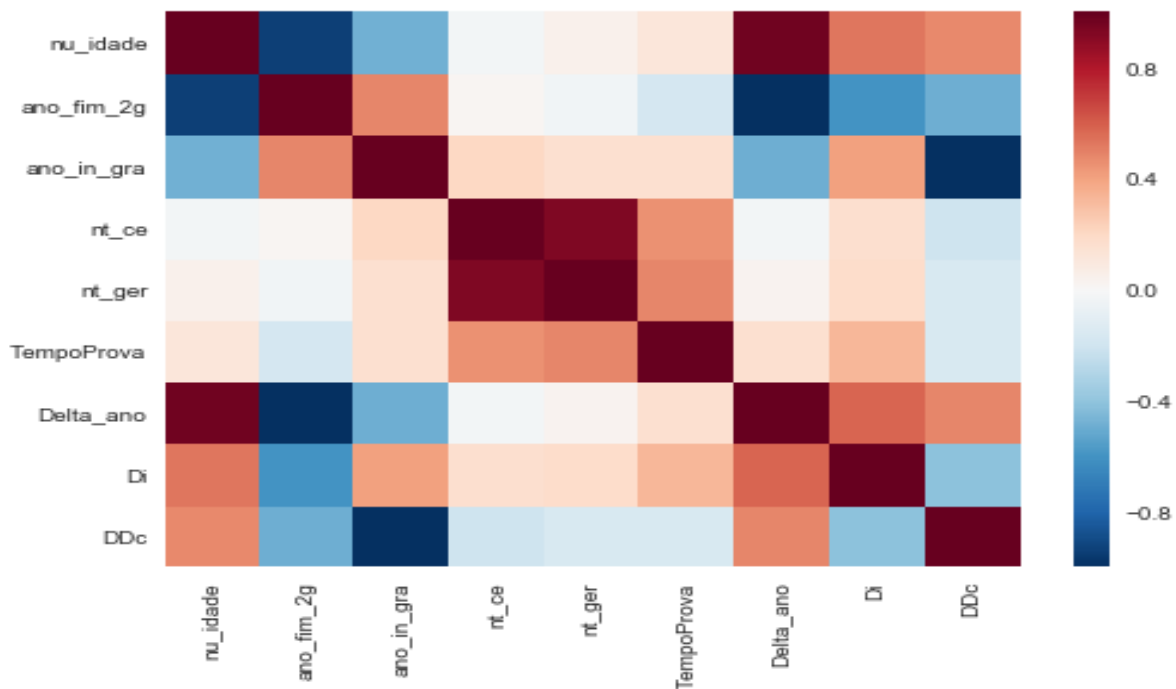
Introdução à Ciência de Dados

Professor: José Viterbo

Autor: Jorge Felipe Campos Chagas

1. INTRODUÇÃO

Este trabalho é uma continuação do trabalho 1 com base nos resultados que obtive no mesmo. No primeiro trabalho tentei relacionar a idade e o tempo que os alunos levaram no desenvolvimento dos seus estudos com seus respectivos desempenhos, e vi que essas variáveis tinham uma relação fraca conforme a figura abaixo (nu_idade, Di, DDC as variáveis temporais).



O resultado obtido foi para o ano de 2011; neste, vou analisar o desempenho dos estudantes ao longo de todos os ENADEs e comparar o desempenho por nota (nota específica vs nota geral) com o desempenho geral ao longo dos anos utilizando regressão linear e alguns dados externos para enfatizar o trabalho.

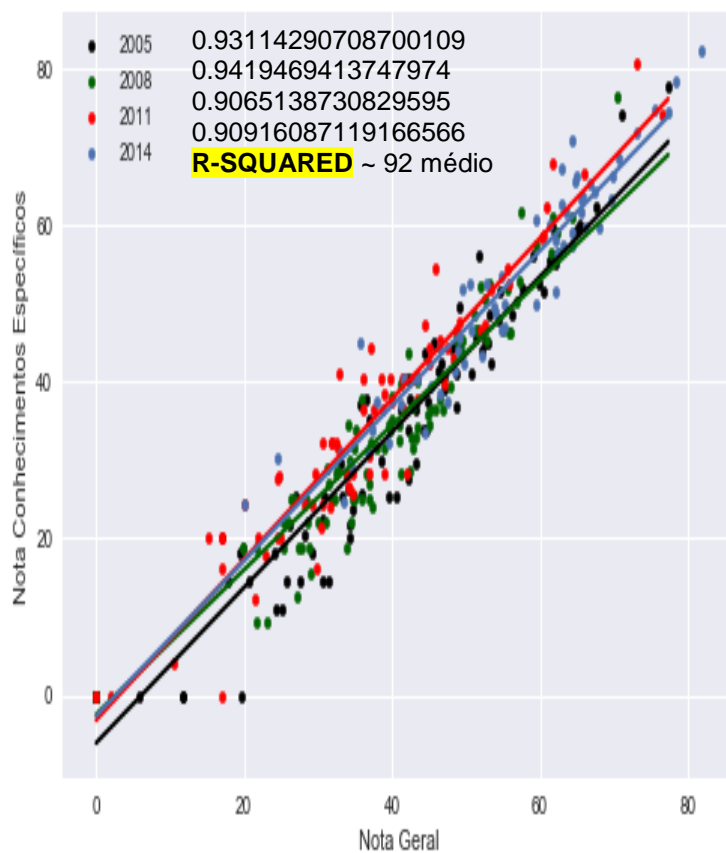
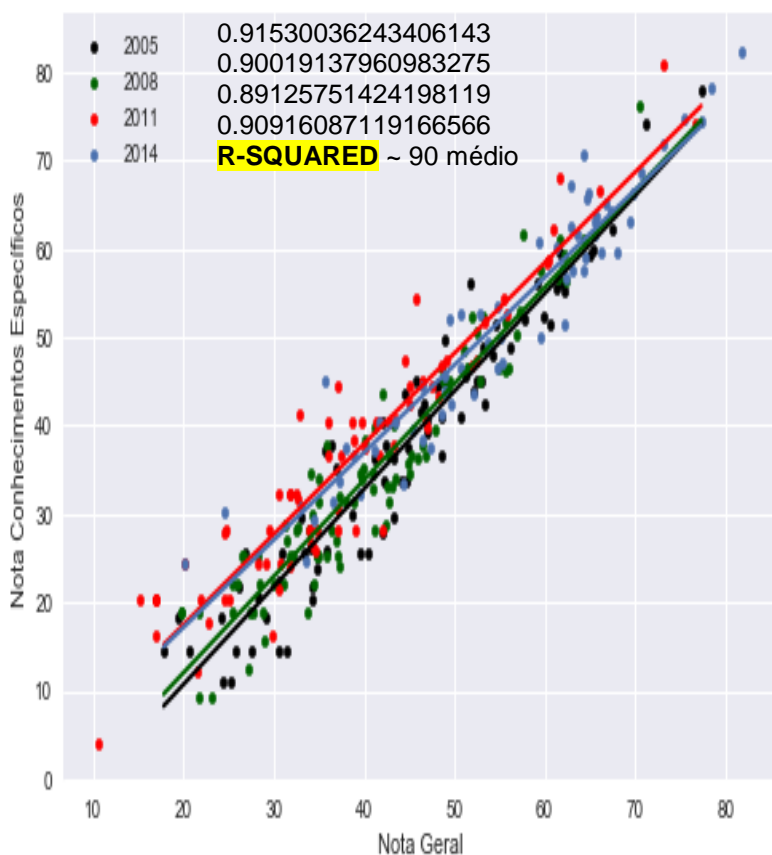
2. METODOLOGIA

Consultar os arquivos (.csv) com os microdados do ENADE (2005 a 2014) e fazer um filtro para selecionar somente os atributos desejados, que são idade, nota geral e nota de conhecimentos específicos e realizar o tratamento dos dados através da linguagem python com o auxílio das bibliotecas numpy, panda, matplotlib e statsmodels.formula.api, as quais ajudaram na construção dos dados visuais e do algoritmo do trabalho. Em seguida, fazer a análise das notas e criar uma projeção dos futuros resultados comparando com o

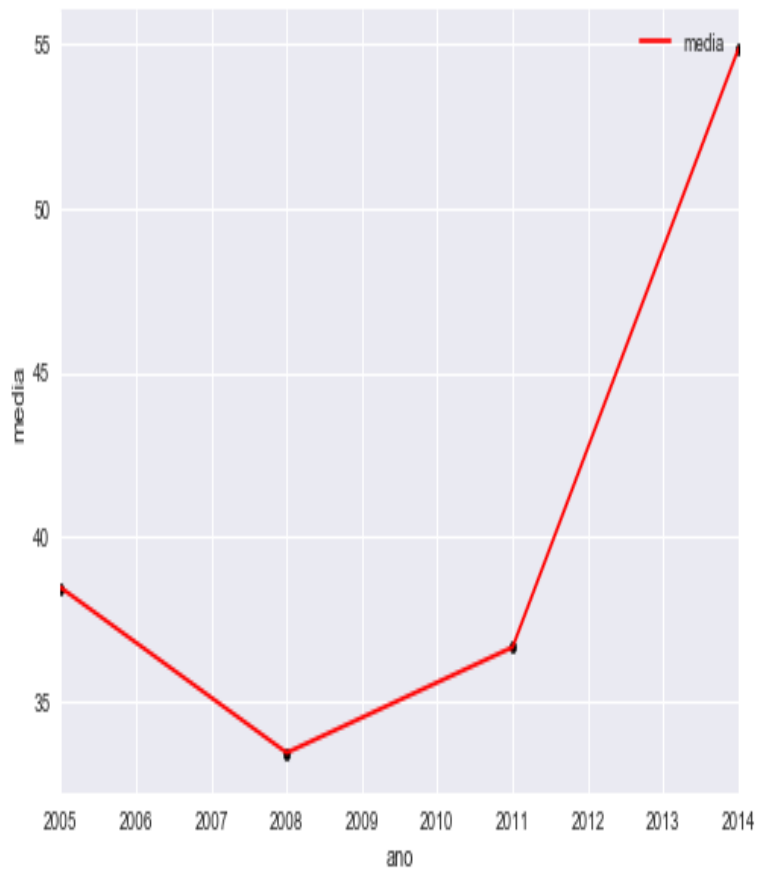
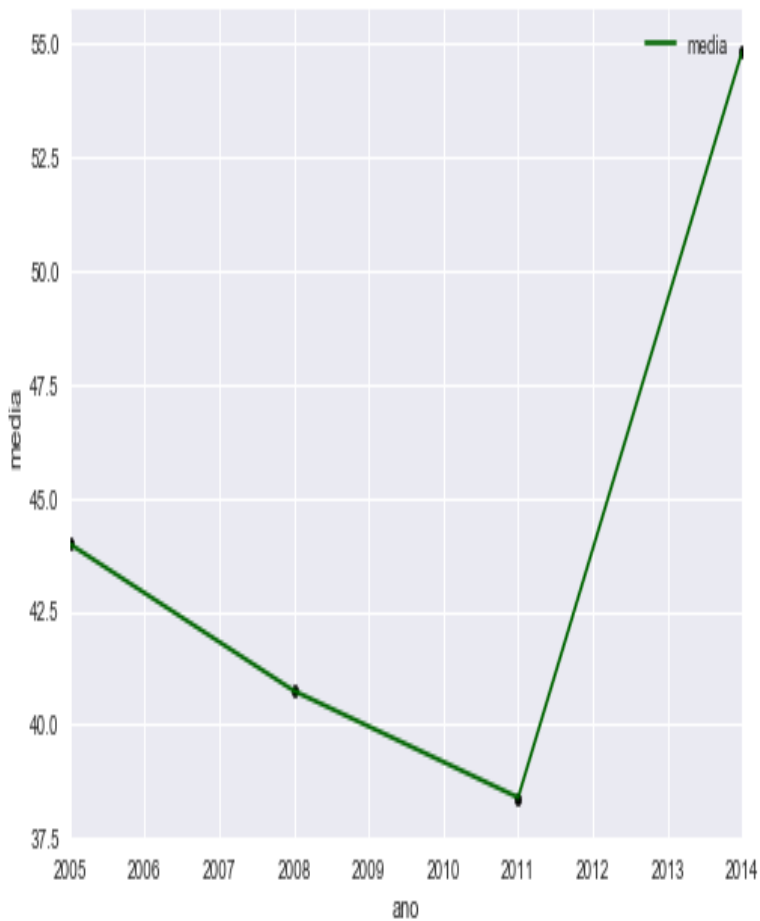
desempenho de estudantes da educação básica em anos posteriores ao ENADE analisado.

3. CÓDIGO e ANÁLISE DE DADOS

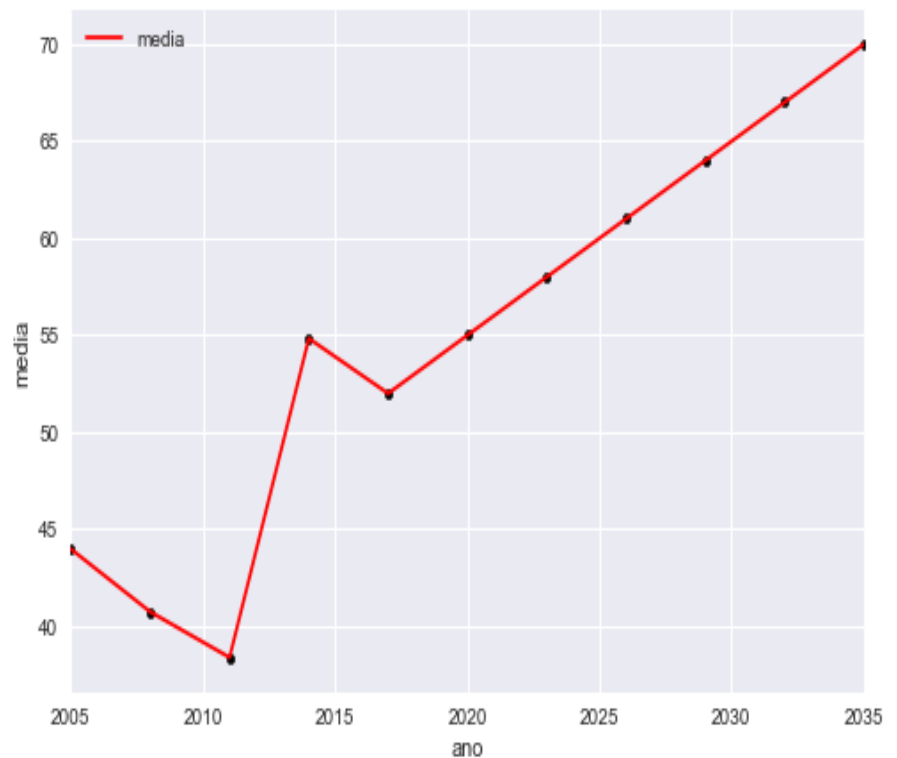
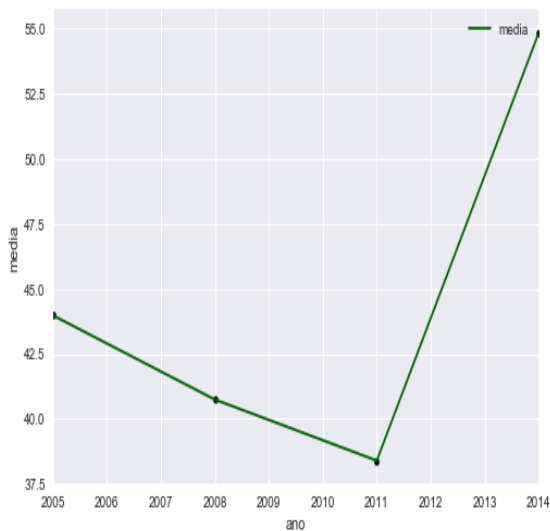
Na primeira parte do código, foi basicamente pegar o CSV e escolher os dados de trabalho. O único tratamento foi dispensar os alunos que obtiveram nota zero ou que faltaram ao exame. Com os dados prontos, fui para a análise, começando pela relação entre nota geral e nota conhecimentos específicos, e obtive o seguinte resultado:



Do lado esquerdo é o gráfico sem considerar os faltosos e as notas zero, do lado direito é o modelo já tratado. Fiquei com o modelo “menos preciso” (maior o RSQUARED maior a confiabilidade do modelo) do lado esquerdo por apresentar uma projeção mais realista ao trabalho, pois no próximo gráfico já teria uma diferença na análise, já que quando fui analisar o desempenho ao longo dos anos no modelo com notas zero o desempenho dos alunos começa a melhorar em 2008, o que não é verdade.



A próxima seguinte foi criar a projeção do modelo através da regressão linear. A projeção mostrada em vermelho foi feita com base no gráfico verde; o modelo previu uma ligeira queda para o ano de 2017 em relação a 2014, por isso essa diferença no gráfico em vermelho.



Depois eu usei minha ideia de macro e micro variáveis e fui buscar dados externos para comparar com minha amostra. Mais o que é uma macro ou uma micro variável? Defini da seguinte forma:

- Macro – variáveis que englobam características menos personalizadas relacionadas ao(s) sujeito(s) de análise.

Ex.: Avaliação do Índice de Desenvolvimento da Educação básica

- Micro – características mais próximas

Ex.: Reformas Curriculares no curso do IC-UFF

Os dados que busquei para comparar com o modelo foram os resultados do IDEB, porém tem que ser observado que os anos não são equivalentes. O aluno que realizou o ENADE em 2011 provavelmente teve influências do IDEB de 3 a 5 anos atrás, que foi a data em que ele saiu do Ensino Médio. Também busquei alguns fatores qualitativos como reformas curriculares e mudança de infraestrutura no IC-UFF.

Ensino Médio

	IDEB Observado						Metas					
	2005	2007	2009	2011	2013	2015	2007	2009	2011	2013	2015	2021
Total	3.4	3.5	3.6	3.7	3.7	3.7	3.4	3.5	3.7	3.9	4.3	5.2
Dependência Administrativa												
Estadual	3.0	3.2	3.4	3.4	3.4	3.5	3.1	3.2	3.3	3.6	3.9	4.9
Privada	5.6	5.6	5.6	5.7	5.4	5.3	5.6	5.7	5.8	6.0	6.3	7.0
Pública	3.1	3.2	3.4	3.4	3.4	3.5	3.1	3.2	3.4	3.6	4.0	4.9

Os resultados marcados em verde referem-se ao Ideb que atingiu a meta.
Fonte: Saeb e Censo Escolar.

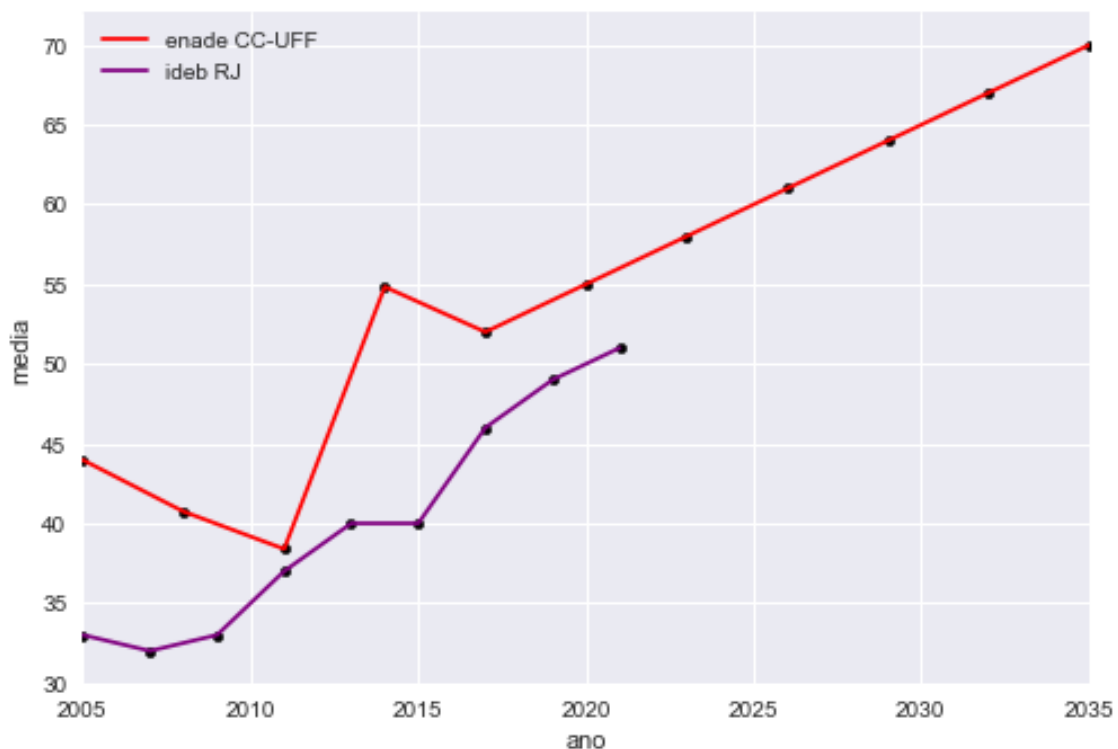
	Ideb Observado						Metas Projetadas							
Estado ↕	2005 ↕	2007 ↕	2009 ↕	2011 ↕	2013 ↕	2015 ↕	2007 ↕	2009 ↕	2011 ↕	2013 ↕	2015 ↕	2017 ↕	2019 ↕	2021 ↕
Rio de Janeiro	3.3	3.2	3.3	3.7	4.0	4.0	3.3	3.4	3.6	3.8	4.2	4.6	4.9	5.1

Obs:

Os resultados marcados em verde referem-se ao Ideb que atingiu a meta.

ano	2005	2008	2011	2014	2017	2020	2023	2026	2029
média	43.97	40.74	38.39	54.83	52.00	55.00	58.00	61.00	64.00

Acima temos na tabela em vermelho o resultado do IDEB ao longo dos anos com suas respectivas projeções (na primeira tabela, o IDEB nacional; na segunda, o IDEB do estado do Rio de Janeiro) e na tabela em azul o ENADE com as expectativas do modelo. Abaixo mostro o gráfico com as projeções do IDEB e do ENADE para fazer um comparativo se minha análise corrobora com a análise feita por especialistas dos futuros resultados do IDEB.



De acordo com o gráfico, olhando para períodos antecessores ao ENADE de análise, vemos que as projeções apontam para expectativas de melhora tanto no desempenho dos alunos do curso de computação da UFF como no dos estudantes do ensino médio do estado do Rio de Janeiro (obs.: assumi que a maioria dos alunos são do estado). Um ponto que vale ressaltar é que o IDEB avalia português e matemática, então poderíamos olhar só as notas de matemática ou fazer uma correlação do desempenho em matemática com o desempenho em conhecimentos específicos e comparar a nota de português com a nota geral do ENADE.

4. CONSIDERAÇÕES FINAIS

Como variáveis externas podem impactar nas projeções do modelo? Por exemplo, corte de verba na área de Educação, escassez de professores no Ensino Médio, a nova reforma curricular que acaba com a obrigatoriedade de algumas disciplinas etc. Outra curiosidade foi a de pegar um conjunto de escolas específicas (de preferência com perfis divergentes – Cap-UFRJ , uma escola particular e uma escola estadual) e ver o desempenho dessas escolas ao longo do tempo. Depois, analisar o que a escola X tem que a faz ter um melhor desempenho que a escola Y. E se pegarmos outra escola qualquer no perfil de X e outra no perfil de Y? Os desempenhos são similares? E assim criar um padrão de desempenho e analisar os possíveis “erros “ que estão levando ao mau desempenho.

5. REFERÊNCIAS

Femi Anthony – Mastering Pandas 2015

<https://pandas.pydata.org/>

<http://dados.gov.br/dataset/microdados-do-exame-nacional-de-desempenho-de-estudantes-ENADE>

<https://stackoverflow.com/questions/>

Roxy Peck, “Statistics: Learning from Data” 2014

<http://IDEB.inep.gov.br/resultado/resultado/resultadoBrasil.seam>