

# Integridade dos Dados de Treinamento para Funcionários Públicos Federais no Brasil

Flávio Garcia Praciano, Bruno J. G Praciano, Fábio L. L de Mendonça  
Erica Lima Gallindo, Daniel Alves da Silva, Francisco C. M Duarte Jr. and Rafael T. de Sousa Jr  
Department of Electrical Engineering, University of Brasília (UnB), Brasília-DF, Brazil  
National School of Public Administration (Enap), Brasília, Brazil  
Email: flavio.praciano@redes.unb.br

**Resumo** — A Escola Virtual de Governo: EV.g vinculada à Escola Nacional de Administração Pública (ENAP) é uma instituição que visa proporcionar capacitação para servidores públicos nas três esferas do Governo, além de formação profissional para cargos de carreira. Tem como objetivo discutir a formação de servidores públicos e outros. A pesquisa bibliográfica qualitativa foi realizada para fins descritivos, utilizando um painel EmNumeros construído com a ferramenta Tableau e o programa de mineração de dados da Python. Para o andamento deste artigo, destacaram a importância do EV.g, atuando como instrumentos de valorização do serviço público, proporcionando a qualificação e o conhecimento desses profissionais das três esferas de governo. têm papel fundamental no andamento da atuação dos servidores públicos, o que representa uma evolução dos serviços para a comunidade e, conseqüentemente, aumenta a qualidade dos serviços oferecidos pela Administração Pública. Conseqüentemente, pode ser comprovado pelos números informados neste artigo que o EV. g descreve-o como um meio de validar de forma confiável a formação de funcionários públicos, visando a construção de servidores públicos ativos e responsáveis em seu papel nas Administrações Públicas.

**Keywords** - *EmNumeros, Enap, Ev.g, Treinamento, Servidores Públicos, Matrículas, Escolas de Governo.*

## I. INTRODUÇÃO

No Brasil, são aproximadamente 11,4 milhões de funcionários públicos. O número total passou de aproximadamente 5,1 milhões para 11,4 milhões de 1986 a 2019. Divididos entre as três esferas de governo (União, Estados, Distrito Federal e Municípios), nos três poderes (Executivo, Legislativo e Judiciário, sendo esses três tipos: (servidores públicos, servidores públicos e funcionários temporários). Esses dados estão disponíveis no painel Em Numeros. Além disso, um portal de cursos oferecidos pelo Governo Federal por meio da Escola Virtual de Governo (EV.g) vinculado à Escola Nacional de Administração Pública (ENAP) com diversos cursos disponibilizados aos servidores públicos de todas as expectativas e comunidades. Essa disponibilidade são dados precisos sobre as qualificações adquiridas por esses públicos [1].

Nesse contexto, o Governo Federal, por meio da EV.g, concentra um conjunto de informações sobre os cursos de capacitação realizados, o que pode despertar o interesse público, como áreas temáticas, cursos realizados, demanda e perfil público para a formação oferecida, agências que utilizam esses cursos de capacitação, número de funcionários treinados,

entre outros. Assim, quando os esforços para fazer isso informação disponível publicamente, controle social é incentivado e a análise de informações de acordo com as mais variadas necessidades. Até agora, essa base unificada armazena cerca de 4,3 milhões de registros realizados entre 2006 e setembro de 2021 [2].

Por meio desse portal, desde 2017, a EV.g tem utilizado a cultura da transparência ativa sobre o serviço prestado, disponibilizando suas informações publicamente sem a necessidade de solicitação prévia dos interessados [2, 4].

A EV.g adotou a cultura da transparência através de seu portal, que fornece acesso aos dados sobre treinamentos oferecidos, permitindo que as partes interessadas façam sua análise das informações fornecidas e fomentando o controle social [4].

A disponibilidade de informações sobre treinamentos realizados no âmbito da EV.g atende à Lei de Acesso à Informação (LAI) [4], que estabelece como dever dos órgãos públicos promover a divulgação de informações de interesse público, independentemente de solicitações, para que a cultura de controle social e transparência estabelecida nos órgãos públicos.

O restante deste artigo é organizado da seguinte forma. A Seção II apresenta os trabalhos relacionados. A seção III descreve o método de análise de dados. Na Seção IV, os resultados são apresentados e discutidos. Finalmente, a Seção V conclui o papel.

## II. OBRAS RELACIONADAS

Os autores de [5] realizaram um estudo dos dados relacionados às informações presentes nos canais de comunicação da ENAP. Embora o trabalho referenciado verifique os dados de apenas um curso, neste artigo, a proposta indica todas as possibilidades de informações a partir dos dados provenientes das bases de dados.

Em Ferrarezi et al. [6], destacou-se a criação de um sistema para registrar o histórico de cursos realizados por cada servidor público e um único cadastro com possibilidades de unificar todas as informações dos usuários em um único local. No entanto, este artigo não criou nenhum sistema, mas criamos um painel Em Numeros para visualização de gerenciamento e um código de mineração de dados.

[7] mostra a importância da formação para os servidores públicos, mas não mostra números reais, e o objetivo deste

artigo é a mineração de dados para uma visualização de entendimento para os gestores para diversos incrementos de novas ações, evidenciado por um painel chamado Em Numeros.

Os autores de [8] pesquisaram aproximadamente 1923 servidores públicos que não completaram um ou mais cursos a distância desenvolvidos pela ENAP, mais significativamente a evasão nos cursos a distância, e um terço deles respondeu à pesquisa por e-mail. Uma ferramenta de coleta de dados do Pacote Estatístico de Ciências Sociais (SPSS) [9] foi utilizada para análise estatística a partir das informações adquiridas na pesquisa enviada aos servidores públicos, e os resultados foram os seguintes: as mulheres (53,9%) abandonaram mais os cursos a distância, principalmente de cursos sem tutoria (54,8%). Em contrapartida, os homens abandonam mais dos cursos com tutoria (52,3%). Neste artigo, a proposta é muito semelhante na qual diz respeito à informação de dados para tomada de decisão. Além disso, o próprio artigo relata a mineração de dados abertos disponíveis para qualquer informação sobre a estática dos cursos e as informações do aluno sobre os cursos oferecidos a eles.

Em [10], os autores fizeram pesquisas exploratórias por meio de uma pesquisa com a ajuda da ENAP. Analisaram diversas distorções de incentivo aos Servidores Públicos Federais, onde se encontrava uma discrepância salarial entre carreiras com o poder executivo e os três poderes com responsabilidades semelhantes. Este artigo citou e evidenciou que há desigualdade salarial entre as carreiras, mas os autores não lidam com alguma formação ou mesmo um avanço na promoção de incentivos de cursos para esses profissionais do Ministério Público Federal, mas este artigo mostra números reais de quantos funcionários estão realizando treinamentos e de quais esferas eles fazem parte.

[11] tem uma proposta de um plugin automático para encorajar mensagens e otimismo para a plataforma Moodle. Essa proposta visa reduzir as taxas de abandono dos alunos em cursos cadastrados por eles. A proposta também visa reduzir as taxas dessas desistências em cursos oferecidos pela EV.g. Segue-se a proposta deste artigo. Diferentemente, o citado apresenta um plugin de mensagem no Moodle, e este artigo analisa e reporta em números reais após o processamento da mineração de dados da plataforma Moodle da Escola Nacional de Administração Pública (ENAP). Esses resultados estão disponíveis em um painel chamado Em numeros: <https://emnumeros.escolavirtual.gov.br/> para decisões de gestão dentro da ENAP.

### III. MATERIAL E MÉTADOS

Esta seção detalha os métodos para verificar a integridade dos dados e as estende à abordagem de aprendizado de máquina. Como mostrado na Figura 1, a análise é dividida em

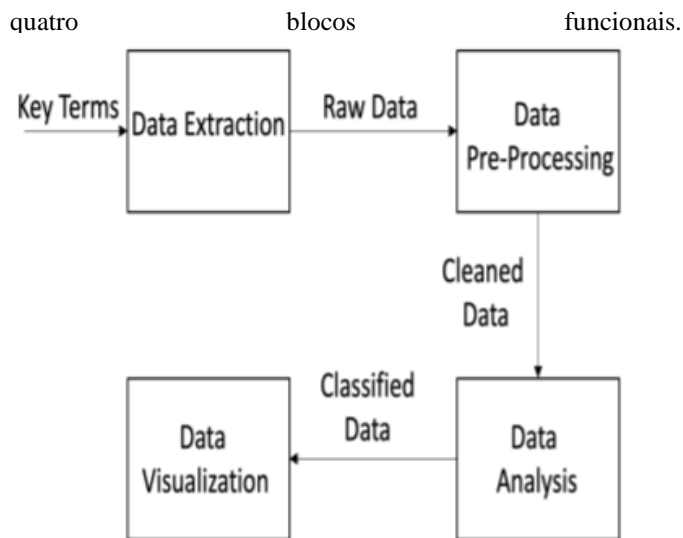


Figure 1: Bloqueie o diagrama da análise proposta para integridade de dados do Portal Em Numeros

#### A. Extração de dados

É possível visualizar as cidades de origem dos matriculados nos cursos de EV.G. que dão cobertura territorial nas regiões Nordeste, Sudeste e Sul, destacando o grande volume de alunos matriculados vindos das capitais desses estados. Ainda assim, no painel indicadores, como primeiro teste do perfil público para os cursos oferecidos no âmbito do EV.g, buscamos identificar o nível de recorrência de matrículas, qual é o percentual de indivíduos que frequentam cursos EV.g mais de uma vez.

O painel apresenta um conjunto de filtros que dão flexibilidade ao usuário para realizar suas análises, permitindo filtrar os gráficos por ano de inscrição (2006 a 2021), a esfera de governo (federal, estadual ou municipal), a esfera de poder (executivo, judiciário ou legislativo), tema, curso e oferta (classe). A reestruturação dos painéis reorganizou as informações que empregam visualizações mais dinâmicas das matrículas e cursos realizados no âmbito da EV.g. Além disso, com base no feedback sobre as informações publicadas, o portal poderá proporcionar melhorias nos serviços prestados, apresentando visualmente informações que possam ser de interesse tanto para a gestão da EV.g quanto para qualquer cidadão comum.

#### B. Pré-processamento de dados

O pré-processamento de dados usou Pandas, onde removemos os dados inconsistentes e os outliers presentes em alguns cursos. É essencial lembrar que os dados pessoais apresentados no conjunto de dados foram removidos. Em relação aos outliers, utilizamos a abordagem Boxplot, onde podemos ver as faixas medianas e interquartil em uma boa forma de visualização. A Figura 2 mostra um exemplo dos outliers da idade dos alunos

matriculados.

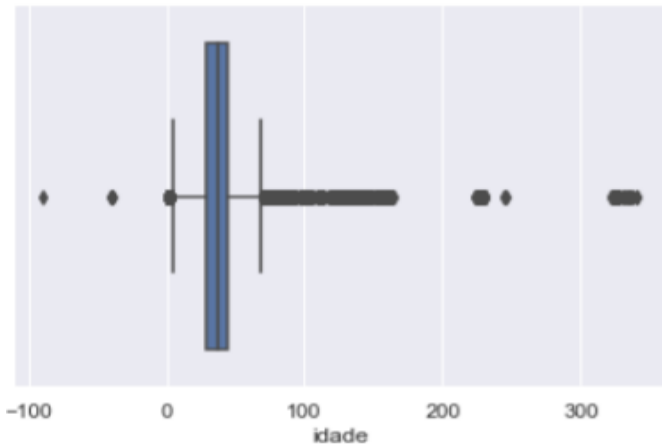


Figure 2: Boxplot com os dados antes do pré-processamento de dados

Usamos a métrica Z-Score para esta remoção outlier que remove o valor da observação acima do valor médio [12]. É necessário realizá-lo porque evitamos o excesso de condições em nosso estimador.

### C. Análise de dados

Propomos uma ferramenta que possa prever o número de alunos que concluirão o curso. Usamos algoritmos de aprendizado de máquina supervisionados depois de rotularmos dados. Além disso, escolhemos algoritmos de regressão para esta tarefa. A análise utiliza o Algoritmo da Floresta Aleatória [13], e para encontrar a quantidade correta de estimadores foi computada com uma matriz de [4, 8, 16, 32, 64, 128, 256].

Além do algoritmo de aprendizado de máquina mencionado, decidimos comparar máquinas vetoriais de suporte (SVM) e uma Rede Neural Artificial (ANN) de duas camadas ocultas. Depois disso, conseguimos os resultados do treinamento para as quatro classes que definimos para a primeira interação (ve, Completo, Falha, Travado e Não Concluído). Para a segunda etapa de treinamento, definimos a fusão das classes em apenas dois aprovados, que contêm o status completo, e outra classe não aprovada com o rótulo de abandono, falha, bloqueio e não preenchido. Uma vez que o conjunto de dados contém mais de 27 colunas, é necessário entender como as principais características estão correlacionadas.

A ANN proposta para este trabalho foi dividida em quatro camadas, onde a camada de entrada contém cinco perceptrons. A Figura 3 mostra a PROPOSTA ANN. Neste caso, uma vez que temos uma classificação binária (Aprovado/Falho), podemos simplesmente usar uma camada de saída de perceptron único; Se a saída for menor que 0,5, o aluno é aprovado; caso contrário, o aluno falha.

Para cada camada, temos como entrada uma matriz feita por colunas de características (em nosso exemplo, temos quatro características, ou seja, Idade do Estudante, Duração do Curso, Sexo Estudantil e Carga horária do curso), que rotulamos como  $I=1, 2, 3, 4$ . Cada um desses recursos terá  $n$  entradas cada recurso é um vetor. A camada  $i$  will have  $p$  perceptrons, labeled

by  $a=1, \dots, p$ . Assim, a saída de toda a camada é uma matriz  $O_{(I)}^{(a)}$  given by:

$$O_{(I)}^{(a)} = \varphi(\bar{w}_{(I)}^{(a)} \cdot x_{(I)}^{(a)} + b_{(I)}^{(a)}), \quad (I)$$

where  $\bar{w}$  is the weight vector and  $b$  is the bias.

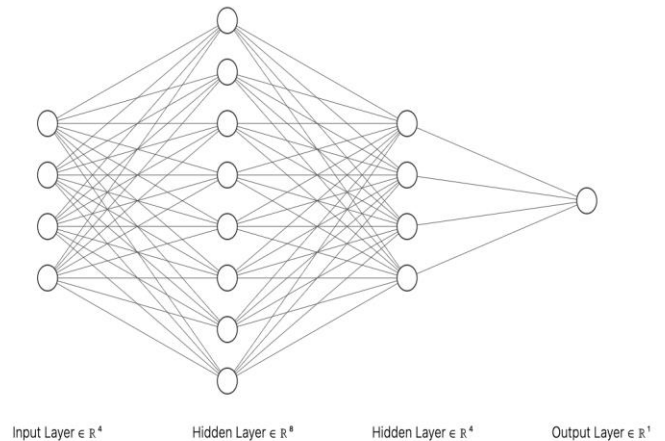


Figure 3: Arquitetura ANN proposta

### D. Visualização de dados

Os dados baixados do Portal Em numeros foram usados para criar alguns gráficos para apoiar os responsáveis por cada curso, seguir as métricas e agir no momento certo. Como o portal mencionado fornece análise visual, precisávamos criar outro tipo de análise visual e uma forma de entender o comportamento dos alunos de alguns cursos. A Figura 4 mostra um painel com os dados disponíveis neste banco de dados aberto.

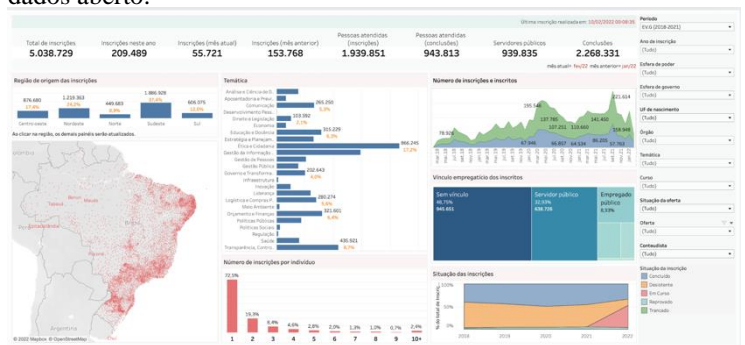


Figure 4: Painel disponível no Portal Em Numeros

## IV. RESULTADOS

Como mencionado na Seção III, usamos três algoritmos diferentes para comparação. Podemos verificar a consistência dos dados de um curso específico porque escolhemos as seguintes características do conjunto de dados: sexo, carga horária do curso, idade do aluno e duração do curso. Esses recursos foram selecionados por meio de uma Análise de Componentes Principais, onde foi possível ver as interações do

recurso. Depois disso, foi possível realizar o treinamento do modelo de aprendizagem de máquina utilizando um número diferente de classificadores. O conjunto de dados foi dividido em treinamento com base em 70% para treinamento, 20% para testes e 10% para validação.

Figure 5 mostra a perda média por época para a Rede Neural.



Figure 5: Average loss by epoch

Decidimos usar a métrica de precisão de cada algoritmo para avaliação, pois temos apenas duas variáveis de objetivo, como Aprovado e Falho. Essa métrica também nos dá uma medida dos pontos de dados relevantes [14] [15]. É importante que não comecemos a tratar um aluno que realmente reprovou como aprovado, mas nosso modelo previu como tê-lo. A Equação II define essa métrica como uma fração dos Verdadeiros Positivos pela soma dos verdadeiros positivos com falsos positivos.

$$Precision = \frac{True\ Positive\ (TP)}{True\ Positive\ (TP) + False\ Positive\ (FP)} \quad (II)$$

Após a configuração do nosso teste de ambiente, poderíamos calcular a métrica selecionada para os três algoritmos. Tabela I mostra o desempenho deles. Comparamos um algoritmo de conjunto (Random Forest), algoritmo linear (SVM) e um Algoritmo multi-camada (ANN).

TABLE I: METRICS RESULTS FOR THE PROPOSED MODEL

	Random Forest	Support Vector Machines	Artificial Neural Network
Precision	0.73	0.83	<b>0.91</b>

Como mostrado na Tabela I, o ANN é a melhor solução quando você precisa estimar uma saída com base em duas classes esperadas. A Figura 6 mostra a matriz de confusão para este algoritmo, é outra maneira de ver o desempenho da nossa solução.

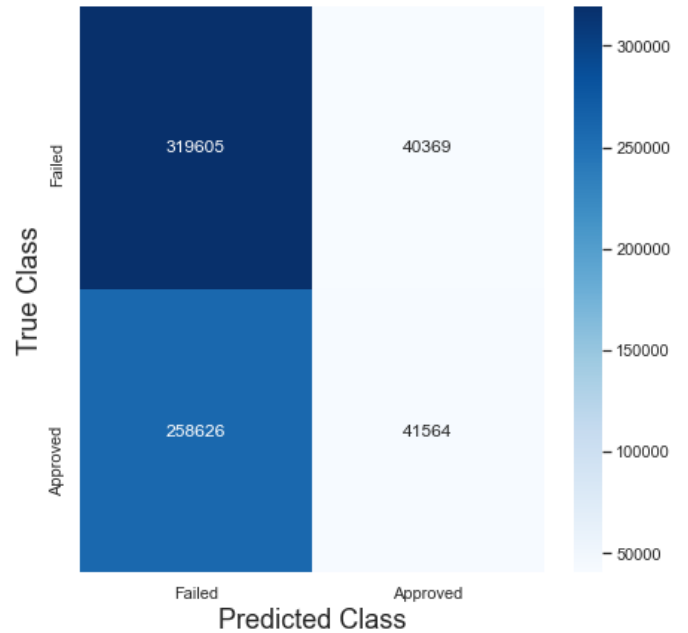


Figure 6: Confusão Matriz da ANN proposta

## V. CONCLUSÕES E OBRAS FUTURAS

Neste trabalho, propomos um quadro para aprovação de um curso com base nos dados do Portal Em Numeros. Nossos resultados, uma precisão próxima de 91% é obtida quando o algoritmo ANN é aplicado para esta tarefa.

Para trabalhos futuros, esperamos criar um pipeline automático de dados, usando uma ferramenta chamada Airflow para automatizar nossa extração de dados e nos permitir melhorar o desempenho de nossa estimativa e apoiar o coordenador do curso para prever o sucesso dos alunos.

## RECONHECIMENTO

Este trabalho é apoiado em parte pelo CNPq – Conselho Nacional de Pesquisas (CNPq). PQ-2 312180/2019-5 sobre Cibersegurança n° 465741/2014-2), em parte pelo Departamento Nacional de Auditoria do Sistema Único de Saúde (SUS). DENASUS 23106.118410/2020-85), em parte pelo Ministério da Economia (MEC). DIPLA 005/2016 e ENAP 083/2016), em parte pelo Conselho Administrativo de Defesa Econômica (N°. CADE 08700.000047/2019-14), em parte pela Procuradoria Geral da União (N°. AGU 697.935/2019), e em parte pela Procuradoria Geral do Tesouro Nacional (N°. PGFN 23106.148934/2019-67) da Fundação de Amparo à Pesquisa do Distrito Federal – FAPDF.

## REFERÊNCIAS

- [1] "Ipea - atlas," in <https://www.ipea.gov.br/atlasestado/consulta/75>, 2019, pp. 1-4.
- [2] "Portal em numeros," in <https://emnumeros.escolavirtual.gov.br/indicadores/>, 2019, pp. 1-4.
- [3] "Escola nacional de administração pública - ENAP," in [www.escolavirtual.gov.br](http://www.escolavirtual.gov.br), 2019, pp. 1-4.
- [4] "Lei nº 12.527, de 18 de novembro de 2011 - lei de acesso a informação LAI," in <http://www.planalto.gov.br/ccivil03/ato2011-2014/2011/lei/112527.htm>, 2011, pp. 1-4.
- [5] V. C. G. Coelho, J. da Costa, D. A. da Silva, R. d. S. Júnior, F. L. de Mendonça, and D. G. Silva, "Mineração de dados educacionais no ensino a distância governamental," Conferências Ibero-Americanas WWW/Internet e Computação Aplicada. Brasília, Brasil, pp. 77-84, 2016.
- [6] E. Ferrarezi and J. A. Tomacheski, "Mapeamento da oferta de capacitação nas escolas de governo no brasil: gestão da informação para fortalecimento da gestão pública," *Revista do Serviço Público*, vol. 61, no. 3, pp. 287-303, 2010.
- [7] A. M. de Andrade, "Escolas de governo e seu papel no aperfeiçoamento do desempenho dos servidores públicos," *RECIMA21-Revista Científica Multidisciplinar-ISSN 2675-6218*, vol. 2, no. 5, pp. e25 350-e25 350, 2021.
- [8] T. P. C. Nascimento and A. Esper, "Evasão em cursos de educação continuada a distância: um estudo na escola nacional de administração pública," *Revista do Serviço Público*, vol. 60, no. 2, pp. 159-173, 2009.
- [9] "Ibm - spss," in <https://www.ibm.com/br-> "Escola nacional de administração pública - ENAP," in [www.escolavirtual.gov.br](http://www.escolavirtual.gov.br), 2019, pp. 1-4.
- [10] I. Corrêa, M. Camões, J. Meyer-Sahling, K. Mikkelsen, and C. Schuster, "Distorções de incentivo ao desempenho e redução de motivação no serviço público federal no brasil," *Revista do Serviço Público*, vol. 71, no. 3, pp. 476-503, 2020.
- [11] L. R. de Almeida, J. P. C. da Costa, R. T. de Sousa, E. P. de Freitas, E. D. Canedo, J. Pretz, E. Zacarias, and G. Del Galdo, "Motivating attendee's participation in distance learning via an automatic messaging plugin for the moodle platform," in 2016 IEEE Frontiers in Education Conference (FIE). IEEE, 2016, pp. 1-5.
- [12] D. Ghosh and A. Vogt, "Outliers: An evaluation of methodologies," in *Joint statistical meetings*, vol. 2012, 2012.
- [13] M. Belgiu and L. Dragut, "Random forest in remote sensing: A review of applications and future directions," *ISPRS journal of photogrammetry and remote sensing*, vol. 114, pp. 24-31, 2016.
- [14] de Oliveira Júnior, G.A.; de Oliveira Albuquerque, R.; Borges de Andrade, C.A.; de Sousa, R.T., Jr.; Sandoval Orozco, A.L.; García Villalba, L.J. Anonymous Real-Time Analytics Monitoring Solution for Decision Making Supported by Sentiment Analysis. *Sensors* 2020, 20, 4557. <https://doi.org/10.3390/s20164557>
- [15] E. S. Gualberto, R. T. De Sousa, T. P. De B. Vieira, J. P. C. L. Da Costa and C. G. Duque, "From Feature Engineering and Topics Models to Enhanced Prediction Rates in Phishing Detection," in *IEEE Access*, vol. 8, pp. 76368-76385, 2020, doi: 10.1109/ACCESS.2020.2989126.