Syahmi Sajid

Portofolio

LIST OF CONTENTS

Introduction

About me

Data science enthusiast with a strong educational background in artificial intelligence Masters at Gadjah Mada University. Mastering Python, SQL, PHP and passion for processing large amounts of data. Has internship experience at Bank Indonesia as a Data Analyst, PT IDX Consulting as a Project-Based Intern: Data Science, Spairum as a Junior Engineer, etc. Have skills and several projects related to EDA, data pre-processing, machine learning, deep learning, computer vision, and NLP. Has several certificates related to Tensorflow and deep learning. Have an interest in a career in data, especially as a data scientist and AI engineer.

Personal **Skill**

Machine Learning

Linier & Logistic Regression, SVM, Random Forest, Kmeans, etc.

Deep Learning

Computer Vision, Natural Language Processing

EDA & Pre-processing Data

Univariate Analysis, Bivariat Analysis, Data Augmentation

Data Visualization App

Tableau and Google Data Studio

Framework Website

Django and Ci4

Microsoft Office

Word, Excel, and PPT

My Experience

Bank Indonesia - Data Analyst (August 2022 - December 2022)

- Mapping and creating QRIS merchant dashboards in West Kalimantan
- Analyzing data on electricity and internet networks and ranking cities for onboarding programs
- Looking for correlations and ranking from BPS and SHP (Data Cental Bank of Indonesia) inflation data per commodity

My Experience

PT IDX Consulting-Project-Based Intern: Data Science Virtual
### (July 2022 – August 2022)

- Finishing various tasks related to the activity of a Data Scientist from PT IDX Consulting, such as Data Analysis, Machine Learning Implementation, and Data Science task

My Experience

## Orbit Future Academy - Artificial Intelligence Mastery - Independent Study (February 2022 - July 2022)

- Studied 4 domains of AI (Data Science, NLP, Computer Vision, Reinforcement Learning)
- Develop an application to help people identify the content of the packaging in a supermarket or minimarket using OCR
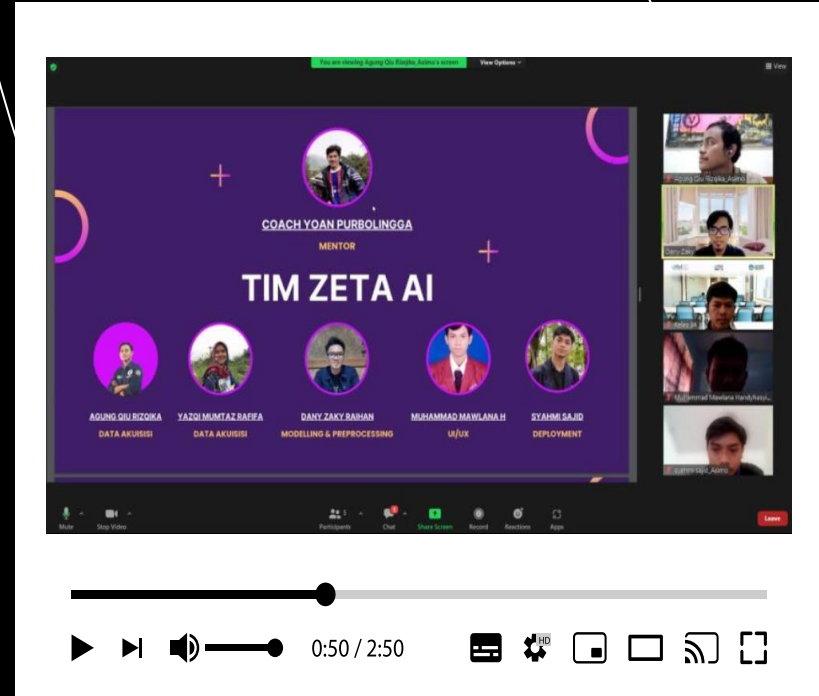
My Experience

Spairum – Junior Engineer (November 2021 - January 2022)

- When apprenticeship results in a user increase of 300%
- Search Insights and visualize 10.000+ row data from available spairum datasets using tableau
- Responsible for managing website backend and 10.000+ data in database using MySQL
- Collaborate with a total of 20 people from the management division and business division to solve business problems

0:50 / 2:50
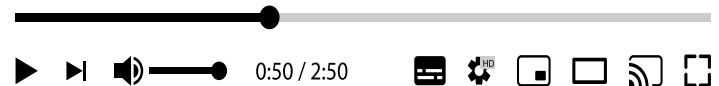
My
Certification

My
Certification

Project

List Project

- Brain Tumor Classification
- Bitcoin Time Series Forecasting with RNN & LSTM
- Credit Fraud with Imbalanced Datasets
- Customer-Personality Segmentation Clustering
- Data Science Salaries (EDA)
- House Price Prediction
- Lung & Colon Cancer Classification
- Skimlit Deep Learning

Project

## Brain Tumor Classification

Description

Proyek ini bertujuan untuk mengembangkan sistem deteksi tumor otak menggunakan model deep learning, khususnya ResNet dan EfficientNet. Sistem ini menggunakan jaringan saraf konvolusional (CNN) untuk mengklasifikasikan gambar MRI otak sebagai positif tumor atau negatif tumor. Data gambar MRI otak dikumpulkan, diproses, dan digunakan untuk melatih model-model CNN tersebut. Tujuannya adalah memilih Model terbaik untuk dipilih dan diimplementasikan sebagai sistem deteksi tumor otak untuk membantu dalam deteksi dini dan perawatan yang tepat.

Data Source

https://www.kaggle.com/datasets/navoneel/brain-mri-images-for-brain-tumor-detection

Link Notebook

https://github.com/Syahmi33github/PortofolioSyahmi/blob/main/Tumor_Classification%20(1).ipynb
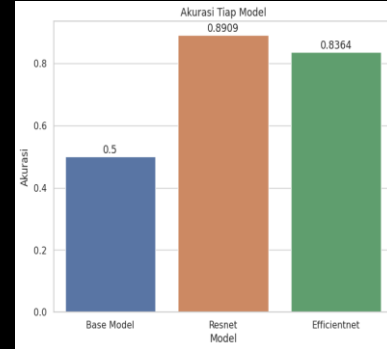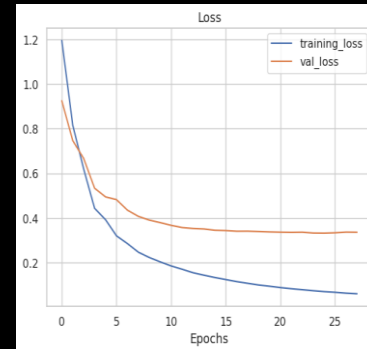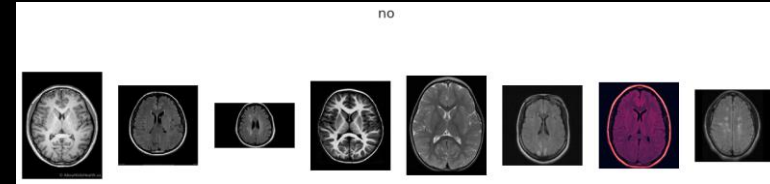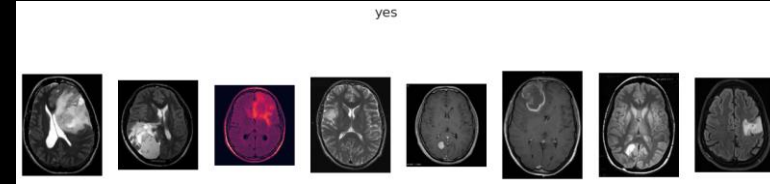
Project

## Bitcoin Time Series Forecasting

Description

Bitcoin adalah cryptocurrency paling lama berjalan dan paling terkenal, pertama kali dirilis sebagai open source pada tahun 2009 oleh Satoshi Nakamoto anonim. Bitcoin berfungsi sebagai media pertukaran digital yang terdesentralisasi, dengan transaksi diverifikasi dan dicatat dalam buku besar yang didistribusikan secara publik (blockchain) tanpa memerlukan otoritas pencatatan tepercaya atau perantara pusat. Blok transaksi berisi hash kriptografi SHA-256 dari blok transaksi sebelumnya, dan dengan demikian "dirantai" bersama, berfungsi sebagai catatan abadi dari semua transaksi yang pernah terjadi. Seperti halnya mata uang/komoditas apa pun di pasar, perdagangan bitcoin dan instrumen keuangan segera mengikuti adopsi bitcoin oleh publik dan terus berkembang. Jika Anda belum mengetahui apa itu Bitcoin, maka dapatkan beberapa pengetahuan tentang Bitcoin di sini.

Data Source

https://www.kaggle.com/datasets/mczielinski/bitcoin-historical-data

Link Notebook

https://github.com/Syahmi33github/PortofolioSyahmi/blob/main/Bitcoin_Time_Series_Forecasting.ipynb
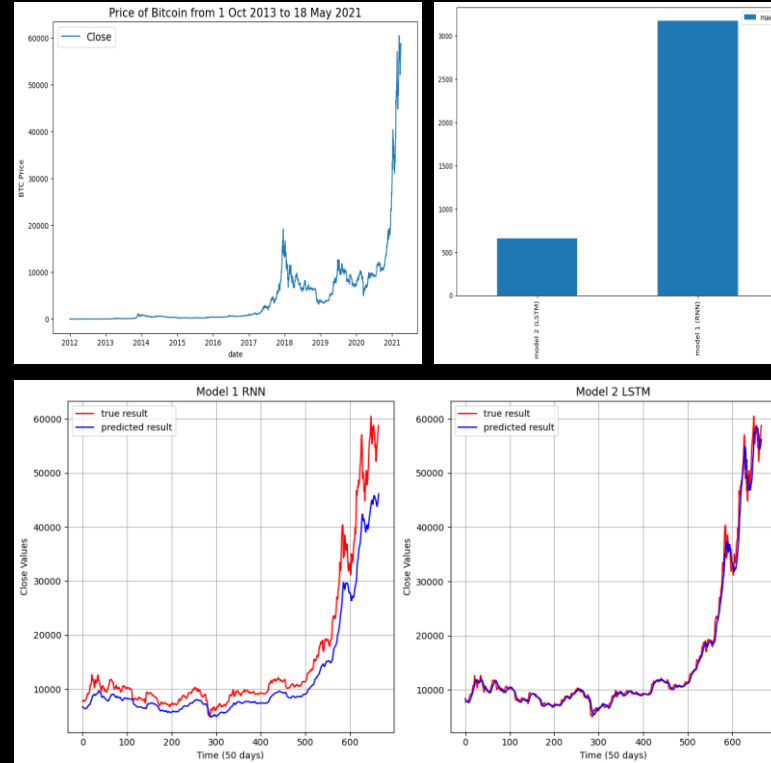
Project

Credit Fraud with Imbalanced Datasets

Description

Project ini bertujuan membuat model terbaik untuk mengklasifikasi credit fraud dengan data yang tidak seimbang. Kumpulan data berisi transaksi yang dilakukan oleh kartu kredit pada bulan September 2013 oleh pemegang kartu Eropa. Dataset ini menyajikan transaksi yang terjadi dalam dua hari, dimana kami memiliki 492 penipuan dari 284.807 transaksi. Dataset sangat tidak seimbang, kelas positif (penipuan) menyumbang 0,172% dari semua transaksi.

Hanya berisi variabel input numerik yang merupakan hasil dari transformasi PCA. Karena masalah kerahasiaan, fitur asli dan lebih banyak informasi latar belakang tentang data tidak disediakan.
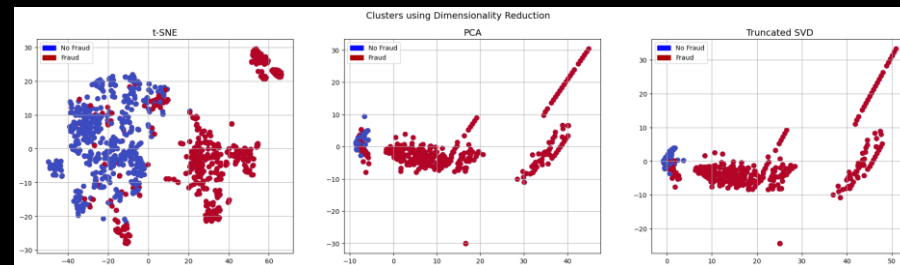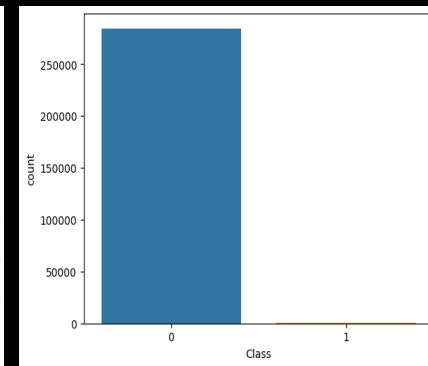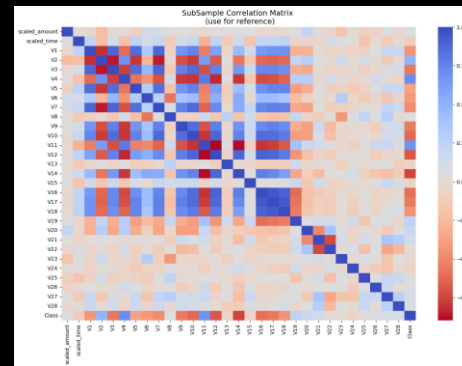Fitur V1, V2, ... V28 adalah komponen utama yang diperoleh dengan PCA;

Data Source

https://www.kaggle.com/datasets/mlg-ulb/creditcardfraud

Link Notebook

https://github.com/Syahmi33github/PortofolioSyahmi/blob/main/Credit_Fraud_with_Imbalanced_Datasets.ipynb

	accuracy
Logistic Regression	93.39%
Support Vector	93.0%
Random Forest	92.74%
KNN	92.73%
DecisionTree	90.75%

Model Terbaik : Logistic Regressiob & SVM

Project

Customer-Personality Segmentation

Description

Analisis Kepribadian Pelanggan adalah analisis terperinci tentang pelanggan ideal perusahaan. Ini membantu bisnis untuk lebih memahami pelanggannya dan mempermudah mereka untuk memodifikasi produk sesuai dengan kebutuhan, perilaku, dan perhatian khusus dari berbagai jenis pelanggan.

Analisis kepribadian pelanggan membantu bisnis untuk memodifikasi produknya berdasarkan target pelanggannya dari berbagai jenis segmen pelanggan. Misalnya, daripada menghabiskan uang untuk memasarkan produk baru ke setiap pelanggan di database perusahaan, perusahaan dapat menganalisis segmen pelanggan mana yang paling mungkin membeli produk dan kemudian memasarkan produk hanya pada segmen tertentu tersebut.

Data Source

https://www.kaggle.com/datasets/imakash3011/customer-personality-analysis

Link Notebook

https://github.com/Syahmi33github/PortofolioSyahmi/blob/main/Customer_Segmentation_Clustering.ipynb
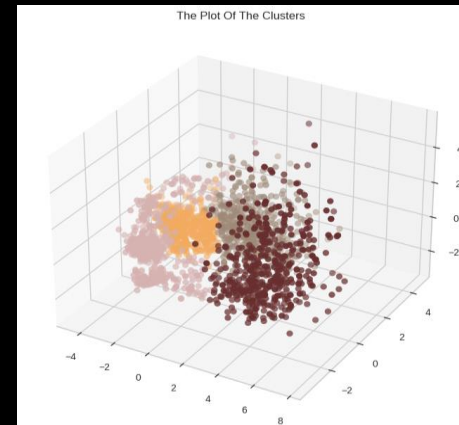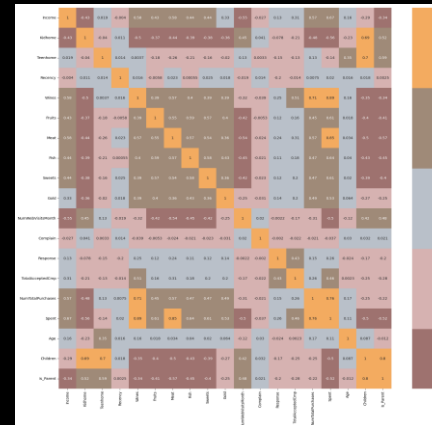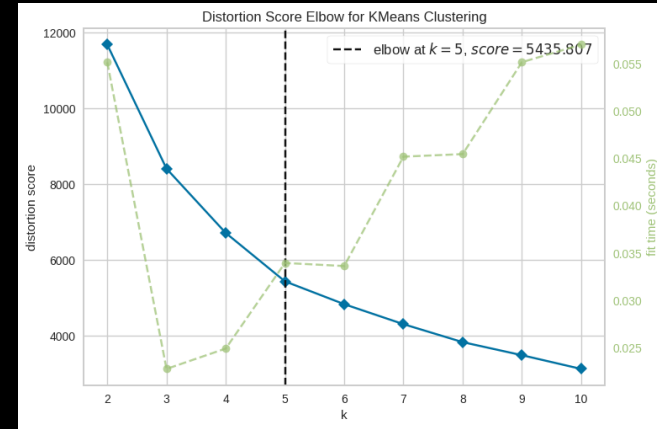
Project

## Data Science Salaries (EDA)

Description

Tujuan dari analisis ini adalah untuk menguji berbagai faktor dan keterkaitannya yang kompleks dengan tujuan untuk memprediksi hasil gaji dengan tepat. Studi ini berusaha untuk memberikan wawasan yang tak ternilai tentang lintasan gaji saat ini di bidang ilmu data, sekaligus meletakkan dasar yang kuat untuk eksplorasi di masa depan tentang dampak faktor penentu utama seperti tingkat pengalaman, jenis pekerjaan, ukuran perusahaan, dan lokasi geografis pada struktur upah. dalam profesi yang berkembang pesat ini.

Data Source

https://www.kaggle.com/datasets/arnabchaki/data-science-salaries-2023

Link Notebook

https://github.com/Syahmi33github/PortofolioSyahmi/blob/main/Data_Science_Salaries_2023.ipynb
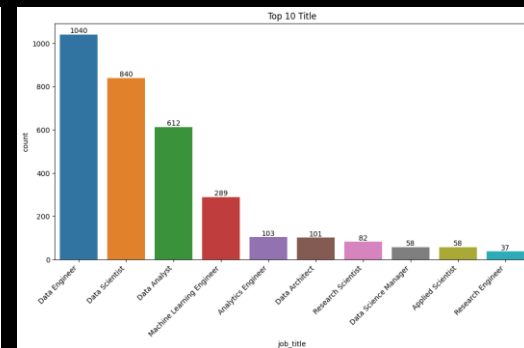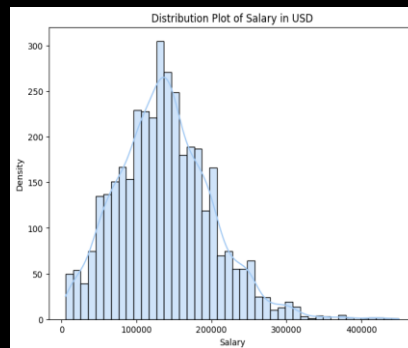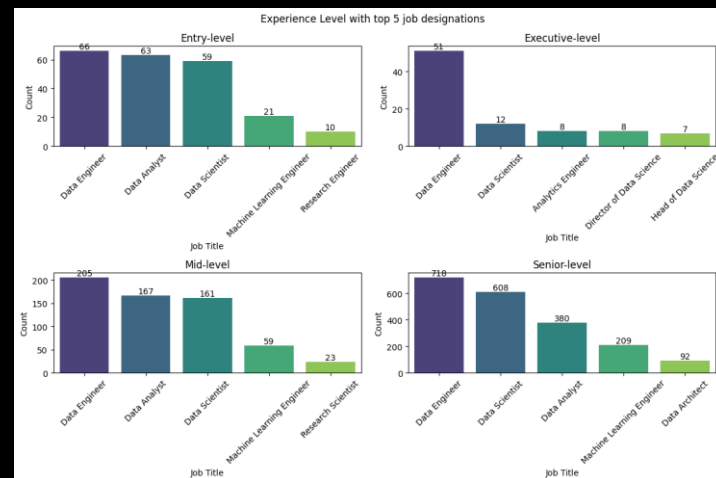
Project

House Price Prediction

Description

Proyek "Prediksi Harga Rumah" bertujuan untuk mengembangkan dan membandingkan kinerja dari tiga model machine learning yang berbeda untuk meramalkan harga-harga perumahan. Model-model tersebut, yaitu Regresi Linier, Mesin Vector Pendukung (SVM), dan Hutan Acak (Random Forest), akan dilatih menggunakan dataset komprehensif yang berisi berbagai fitur yang mempengaruhi harga rumah. Melalui proyek ini, kami bertujuan untuk menilai kemampuan prediksi dari masing-masing model dan menentukan pendekatan yang paling akurat dan sesuai untuk memperkirakan harga rumah berdasarkan data yang diberikan.

Data Source

https://www.kaggle.com/datasets/vikrishnan/boston-house-prices

Link Notebook

https://github.com/Syahmi33github/PortofolioSyahmi/blob/main/House_Price.ipynb

Linier Regression

Random Forest

SVM

	Model	MAE	MSE	RMSE
1	Random Forest	0.091033	0.016013	0.126544
2	Support Vector Machines	0.098861	0.017941	0.133945
0	Linear Regression	0.132017	0.029451	0.171612

Project

## Lung & Colon Cancer Classification

Description

Proyek ini bertujuan untuk mengembangkan sistem klasifikasi kanker paru-paru dan kolon menggunakan model deep learning, terutama Convolutional Neural Network (CNN), ResNet, dan EfficientNet. Dataset yang berisi citra medis seperti gambar X-ray paru-paru dan gambar kolon dikumpulkan, diproses, dan digunakan untuk melatih model-model CNN tersebut. Dengan menggunakan model-model tersebut, proyek ini bertujuan untuk mengklasifikasikan gambar-gambar medis sebagai kanker paru-paru positif atau negatif, serta kanker kolon positif atau negatif. Sistem ini diharapkan dapat membantu dalam deteksi dini kanker paru-paru dan kolon, sehingga memungkinkan penanganan dan pengobatan yang lebih efektif.

Data Source

https://www.kaggle.com/datasets/andrewmvd/lung-and-colon-cancer-histopathological-images

Link Notebook

https://github.com/Syahmi33github/PortofolioSyahmi/blob/main/Lung_%26_Colon_Cancer_Classification.ipynb
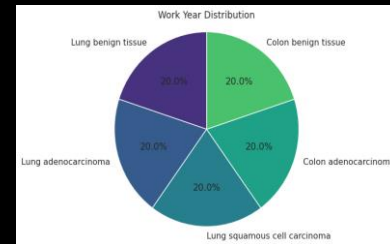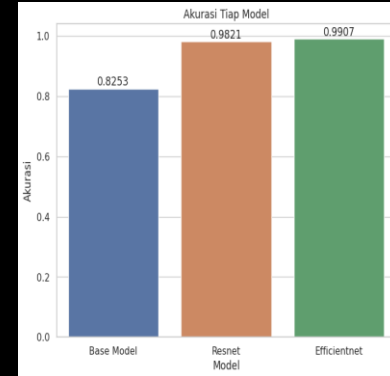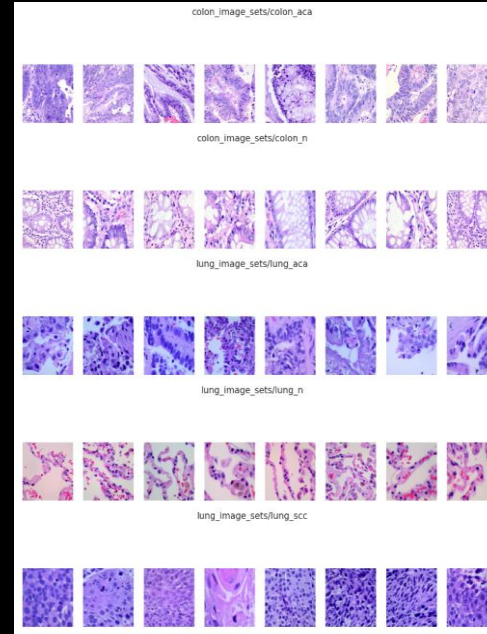
Project

SkimLit NLP

Nutritional psychiatry: the present state of the evidence

Wolfgang Marx [1], Genevieve Moseley [2], Michael Berk [2], Felice Jacka [2]

Affiliations + expand
PMID: 28942748   DOI: 10.1017/S0029665117002026

Abstract

Mental illness, including depression, anxiety and bipolar disorder, accounts for a significant proportion of global disability and poses a substantial social, economic and health burden. Treatment is presently dominated by pharmacotherapy, such as antidepressants, and psychotherapy, such as cognitive behavioural therapy; however, such treatments avert less than half of the disease burden, suggesting that additional strategies are needed to prevent and treat mental disorders. There are now consistent mechanistic, observational and interventional data to suggest diet quality may be a modifiable risk factor for mental illness. This review provides an overview of the nutritional psychiatry field. It includes a discussion of the neurobiological mechanisms likely modulated by diet, the use of dietary and nutraceutical interventions in mental disorders, and recommendations for further research. Potential biological pathways related to mental disorders include inflammation, oxidative stress, the gut microbiome, epigenetic modifications and neuroplasticity. Consistent epidemiological evidence, particularly for depression, suggests an association between measures of diet quality and mental health, across multiple populations and age groups; these do not appear to be explained by other demographic, lifestyle factors or reverse causality. Our recently published intervention trial provides preliminary clinical evidence that dietary interventions in clinically diagnosed populations are feasible and can provide significant clinical benefit. Furthermore, nutraceuticals including n-3 fatty acids, folate, S-adenosylmethionine, N-acetyl cysteine and probiotics, among others, are promising avenues for future research. Continued research is now required to investigate the efficacy of intervention studies in large cohorts and within clinically relevant populations, particularly in patients with schizophrenia, bipolar and anxiety disorders.

Source: https://pubmed.ncbi.nlm.nih.gov/28942748/

Harder to read

Wall of text broken into chunks

Model

SkimLit

Considerations for a surgical RCT for low-grade glioma: a survey

Alireza Mansouri [1], Karanbir Brar [2], Michael D Cusimano [3]

Affiliations + expand
PMID: 32537182   PMCID: PMC7274180 (available on 2021-06-01)   DOI: 10.1093/nop/npz058

Abstract

Background: Diffuse low-grade gliomas (DLGGs) are heterogeneous tumors that inevitably differentiate into malignant entities, leading to disability and death. Recently, a shift toward up-front maximal safe resection of DLGGs has been favored. However, this transition is not supported by randomized controlled trial (RCT) data. Here, we sought to survey the neuro-oncology community on considerations for a surgical RCT for DLGGs.
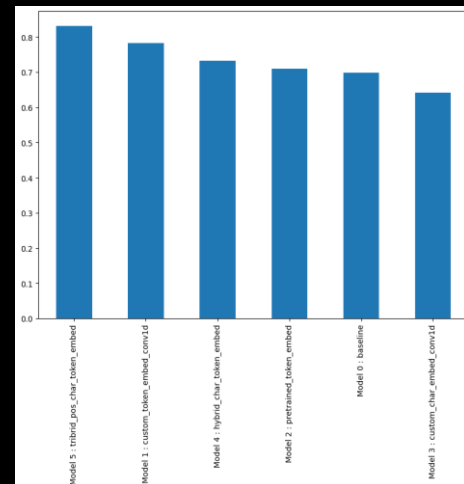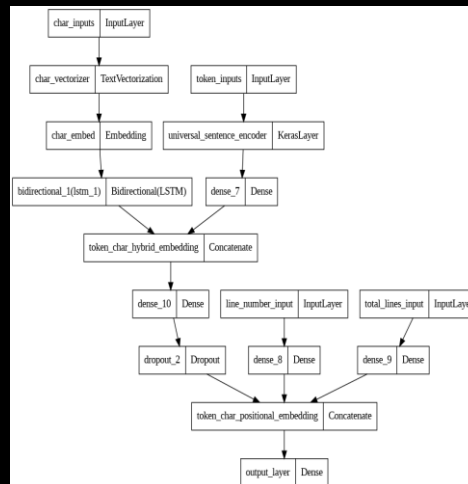
Methods: A 21-question survey focusing on a surgical RCT for DLGGs was developed and validated by 2 neurosurgeons. A sample case of a patient for whom management might be debatable was presented to gather additional insight. The survey was disseminated to members of the Society for Neuro-Oncology (SNO) and responses were collected from March 16 to July 10, 2018.

Results: A total of 131 responses were collected. Sixty-three of 117 (54%) respondents thought an RCT would not be ethical, 39 of 117 (33%) would consider participating, and 56 of 117 (48%) believed an RCT would be valuable for determining the differing roles of biopsy, surgery, and observation. This was exemplified by an evenly distributed selection of the latter management options for our sample case. Eighty-three of 120 (69.2%) respondents did not believe in equipoise for DLGG patients. Quality of life and overall survival were deemed equally important end points for a putative RCT.

Conclusions: Based on our survey, it is evident that management of certain DLGG patients is not well defined and an RCT may be justified. As with any surgical RCT, logistic challenges are anticipated. Robust patient-relevant end points and standardization of perioperative adjuncts are necessary if a surgical RCT is undertaken.

Source: https://pubmed.ncbi.nlm.nih.gov/32537182/

Easier to read
(skimmable)

Data Source

https://github.com/Franck-Dernoncourt/pubmed-rct.git

Link Notebook

https://github.com/Syahmi33github/PortofolioSyahmi/blob/main/SkimLit_NLP%20(1).ipynb

	accuracy	precision	recall	f1
Model 0 : baseline	72.183238	0.718647	0.721832	0.698925
Model 1 : custom_token_embed_conv1d	78.439031	0.780880	0.784390	0.782098
Model 2 : pretrained_token_embed	71.273004	0.712964	0.712730	0.709805
Model 3 : custom_char_embed_conv1d	65.212498	0.645840	0.652125	0.642141
Model 4 : hybrid_char_token_embed	73.417847	0.735862	0.734178	0.732251
Model 5 : tribrid_pos_char_token_embed	83.205349	0.830934	0.832053	0.831318

THANKS