

مطالعه تطبیقی الگوریتم های یادگیری نظارت شده برای پیش بینی عملکرد تحصیلی دانشجویان

A comparative study of Supervised learning algorithms for predicting students' academic performance

مقدمه

بدون شک، مهمترین دارایی دانشگاه ها و موسسات آموزش عالی دانشجویان هستند و این نهادهای آموزش عالی بدون دانشجویان ارزشی ندارند. در واقع، دانشجویان به عنوان مهمترین سرمایه انسانی برای نهادهای آموزش عالی و جامعه در کمک به توسعه همه جانبه یک کشور شناخته می شوند. به ویژه اینکه، در عصر حاضر، با انتقال از اقتصاد مبتنی بر تولید به اقتصاد دانش بنیان به ویژه در کشورهای در حال توسعه این موضوع اهمیت زیادی یافته است (رحمان^۱ و همکاران، ۲۰۲۱). اقتصاد دانش بنیان اقتصادی است که از دانش به عنوان موتور اصلی رشد اقتصادی استفاده می شود. اساساً، این اقتصاد عامل اصلی رشد، ثروت آفرینی و اشتغال در تمامی زمینه ها می باشد (فال^۲ و همکاران، ۲۰۲۱). یکی از عناصر کلیدی برای تحقق اقتصاد دانش بنیان سرمایه های انسانی، به ویژه سرمایه های دانشگاهی (دانشجویان) قلمداد می شود. دال بر این ادعا، مطالعات زیادی تأثیر سرمایه انسانی دانشی را بر رشد اقتصادی بررسی کرده اند که گواه نقش سرمایه های دانشی بر رشد اقتصادی می باشد (نگپاه^۳ و همکاران، ۲۰۲۱). در واقع، توسعه اقتصادی و اجتماعی یک کشور ارتباط مستقیمی با عملکرد تحصیلی دانشجویان دارد. عملکرد تحصیلی دانشجویان نقشی حیاتی در ایجاد فارغ التحصیلان با بهترین کیفیت ایفا می کند که رهبر و نیروی انسانی یک کشور خاص و در نتیجه مسئول توسعه اجتماعی و اقتصادی کشور خواهند شد (علی^۴ و همکاران، ۲۰۰۹). به همین دلیل موضوع عملکرد تحصیلی دانشجویان همواره در تحقیقات گذشته در کشورهای مختلف مورد توجه قابل توجه قرار گرفته است. عملکرد دانشجویان تحت تأثیر عوامل متعددی همچون روانی، اقتصادی، اجتماعی، فردی و محیطی است. اگرچه این عوامل به شدت بر عملکرد دانشجویان تأثیر می گذارند، اما این عوامل از کشوری به کشور دیگر و فردی به فرد دیگر متفاوت است (سینگ^۵ و همکاران، ۲۰۱۶). ما در این پژوهش معدل دانشجویان را معادل عملکرد تحصیلی قرار داده ایم.

معدل نشانگر مهمی برای نمایش توانایی تحصیلی دانشجویان است. ما بر اساس معدل دانشجویان را در دسته های خوب و بد گروه بندی می کنیم. همچنین، معدل نشانگری مهم در تصمیمات بعدی دانشجویان نظیر تصمیم به ادامه تحصیل، تصمیم به درخواست تحصیل در سایر کشورها و غیره است. در بازار کار نیز، معدل به عنوان یک سینگال مهم از کیفیت تحصیلی دانشجو در نزد کارفرمایان است. عوامل بسیار زیادی بر معدل تأثیر گذارند. فلات و پیسکوا^۶ (۲۰۲۲) در پژوهشی به پیش

¹ Rehman

² Phale

³ Ngepah

⁴ Ali

⁵ Singh

⁶ Falát & Piscová

بینی معدل دانشجویان با استفاده از تکنیک های هوش مصنوعی پرداختند و به این عوامل موثری همچون: انگیزه تحصیل، مطالعه منظم در کل ترم، حضور در سخنرانی، استفاده از مشاوره در طول مطالعات، تماشای ویدیوهای آموزشی خاص در یوتیوب، معدل سال اول دانشگاه، معدل دبیرستان، میزان استفاده از شبکه های اجتماعی، تعداد ساعات خواب در طول شبانه روز را شناسایی کردند. خان^۷ و همکاران (۲۰۲۱) نیز در مطالعه ای با استفاده از الگوریتم های داده کاوی برای پیش بینی معدل دانشجویان کارشناسی علوم کامپیوتر، عوامل موثری همچون؛ تعداد ساعات خواب در شب، دفعات مصرف مواد غیرقانونی، تعداد ساعات مطالعه در هفته و تعداد ساعات صرف شده در شبکه های اجتماعی در هفته بر معدل شناسایی کردند. ملانا و دفریانی^۸ (۲۰۲۰) در پیش بینی عوامل موثر بر معدل دانشجویان با استفاده از تکنیک های هوش مصنوعی، عواملی نظیر؛ طول مدت تحصیل دانشجویان معدل موقت نیمسال اول، معدل موقت نیمسال دوم، وضعیت سازمانی و وضعیت اشتغال را مطرح کردند. اله^۹ (۲۰۱۹) با استفاده از الگوریتم های رگرسیون یادگیری ماشین و عمیق برای پیش بینی معدل دانشجویان، عواملی همچون معدل دبیرستان، نتایج آزمون تعیین سطح و نمره آیلتس پیش بینی کننده های مهمی برای معدل تلقی می شوند. پوتپوک^{۱۰} و همکاران (۲۰۱۸) نیز با استفاده از الگوریتم های هوش مصنوعی در شناسایی عوامل تأثیرگذار بر معدل فارغ التحصیلی، عواملی همچون؛ جنسیت دانشجو، بورسیه اعطا شده، سابقه تحصیلی قبلی، نوع پذیرش، استعداد و استان دبیرستان محل تحصیل را ذکر کردند. احمد^{۱۱} و همکاران (۲۰۱۵) عوامل موثری هم مانند اطلاعات دموگرافیک، سوابق تحصیلی قبلی و اطلاعات پیشینه خانوادگی دانشجویان بر معدل را با استفاده از تکنیک های هوش مصنوعی را بیان کردند. همانگونه که تشریح شد، معدل دانشجویان تابعی از عوامل مختلفی است که این عوامل در سیستم آموزش عالی هر کشور متفاوت است. زیرا هر کشور شرایط اقتصادی، اجتماعی، فرهنگی و سیاسی خود را دارد که در سطوح مختلف بر سیستم آموزش عالی آن کشور و به تبع آن بر معدل (عملکرد تحصیلی) دانشجویان اثر گذار است. ما در این مطالعه به پیش بینی معدل دانشجویان دانشگاه تهران می پردازیم. دانشگاه تهران یک دانشگاه دولتی و یکی از بزرگترین مراکز آموزش عالی در ایران است. از این دانشگاه با القاب «دانشگاه مادر» و «نماد آموزش عالی» ایران یاد شده است. با توجه به اهمیت و رشد سریع دستاوردهای هوش مصنوعی در حوزه های مطالعاتی و همچنین دقت بالای آن در پیش بینی به نسبت تحلیل های آماری مرسوم، این پژوهش با هدف، مطالعه تطبیقی الگوریتم های نظارت شده برای پیش بینی عملکرد تحصیلی دانشجویان انجام شده است.

روش شناسی

یادگیری ماشینی به عنوان یک رشته تحقیقاتی در زمینه هوش مصنوعی در سال های اخیر ظهور کرده است. بر اساس این روش یادگیری، بین یادگیری نظارت شده و یادگیری بدون نظارت تمایز قائل می شویم. همانطور که در این تحقیق بر

⁷ Khan

⁸ Maulana & Defriani

⁹ ALLAH

¹ Putpuek

¹ Ahmad

یادگیری تحت نظارت متمرکز شدیم، روش‌های نظارتی را که بعداً برای ایجاد مدل‌های پیش‌بینی استفاده کردیم به اختصار ارائه می‌کنیم.

- مدل‌های مورد استفاده

رگرسیون خطی چندگانه

Linear regression is a commonly used statistical technique for modeling the relationship between a dependent variable and one or more independent variables. It is used to predict the value of the dependent variable based on the values of the independent variables. In linear regression, a linear relationship is assumed between the dependent variable and the independent variable(s), and the aim is to find the best-fitting line that minimizes the sum of the squared differences between the predicted values and the actual values.

Linear regression can be used for both simple and multiple regression models. Simple linear regression involves one independent variable, while multiple linear regression involves two or more independent variables. The technique is widely used in various fields such as finance, economics, engineering, and social sciences for making predictions and understanding relationships between variables. A reference for further reading on linear regression can be found in the book "Applied Linear Regression" by Sanford Weisberg (Wiley, 2014).

Mean Square Error (MSE), Mean Absolute Error (MAE), and Root Mean Square Error (RMSE) are three commonly used metrics to evaluate the performance of a regression model. They are measures of the average difference between the predicted values and the actual values.

MSE is the average of the squared differences between the predicted values and the actual values. It is given by:

$$MSE = \frac{1}{n} \sum (y_{pred} - y_{actual})^2$$

where y_{pred} is the predicted value, y_{actual} is the actual value, and n is the number of data points.

MAE is the average of the absolute differences between the predicted values and the actual values. It is given by:

$$MAE = \frac{1}{n} \sum |y_{pred} - y_{actual}|$$

where $|x|$ represents the absolute value of x .

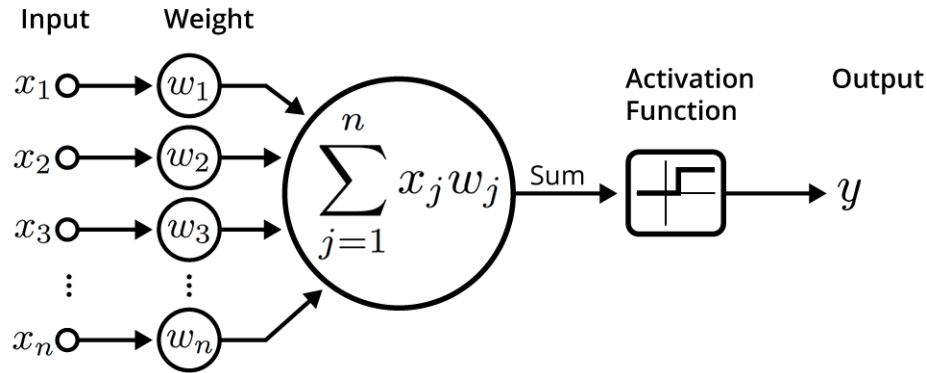
RMSE is the square root of the MSE. It is given by:

$$RMSE = \sqrt{MSE}$$

The RMSE is often used as a preferred metric because it has the same unit as the dependent variable and is more interpretable than the MSE.

In general, a lower value of these metrics indicates better performance of the model. However, the choice of metric depends on the specific problem and the goals of the analysis.

رگرسیون شبکه عصبی



Neural networks are a class of machine learning algorithms that are inspired by the structure and function of the human brain. They consist of interconnected nodes, called neurons, that work together to perform complex computations. Neural networks are capable of learning patterns and relationships in data, making them suitable for a wide range of applications such as image and speech recognition, natural language processing, and autonomous control. The ability to learn from data makes neural networks useful in situations where traditional rule-based systems would be impractical or impossible to implement.

Neural networks can be trained using various algorithms, including backpropagation and stochastic gradient descent. The training process involves adjusting the weights and biases of the neurons in the network to minimize the error between the predicted output and the actual output. A reference for further reading on neural networks can be found in the book "Deep Learning" by Ian Goodfellow, Yoshua Bengio, and Aaron Courville (MIT Press, 2016).

Neural network structure is optimized in this analysis via hyperparameter tuning to have more precise prediction. Hyperparameter tuning is a critical step in building neural network models that involves selecting the optimal values for the hyperparameters. Hyperparameters are parameters that are set before the training process begins, and they affect the behavior and performance of the neural network.

Common hyperparameters that require tuning include the learning rate, number of hidden layers, number of neurons in each layer, activation function, and regularization strength. The optimal values for these hyperparameters depend on the specific problem, dataset, and architecture of the neural network. Hyperparameter tuning can be performed using various methods, including grid search, random search, Bayesian optimization, and genetic algorithms. These methods involve searching the hyperparameter space to find the optimal combination of hyperparameters that maximize the performance of the neural network on a validation set.

A reference for further reading on hyperparameter tuning in neural networks can be found in the book "Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow" by Aurélien Géron (O'Reilly Media, 2019).

الگوریتم جنگل تصادفی

Random forests are a popular machine learning algorithm that is used for classification and regression tasks. They are an ensemble learning method that combines multiple decision trees to improve the accuracy and robustness of the model. In a random forest, multiple decision trees are trained on random subsets of the data and random subsets of the features. This randomness helps to reduce overfitting and improve the generalization ability of the model. The final prediction is made by averaging the predictions of all the trees in the forest.

Random forests are widely used in various fields such as finance, healthcare, and image analysis, where they have shown to be effective in handling high-dimensional data and dealing with missing or noisy data. A reference for further reading on random forests can be found in the paper "Random Forests" by Leo Breiman (Machine Learning, 2001).

الگوریتم درخت تصمیم

Decision trees are a popular machine learning algorithm that is used for classification and regression tasks. They are a type of supervised learning algorithm that works by recursively partitioning the data into subsets based on the values of the input features, until a stopping criterion is met. In a decision tree, each internal node represents a test on a feature, each branch represents the outcome of the test, and each leaf node represents a class label or a numerical value. The structure of the tree is learned from the training data using various algorithms, such as the ID3, C4.5, or CART algorithm.

Decision trees are widely used in various fields such as finance, healthcare, and image analysis, where they have shown to be effective in handling high-dimensional data and dealing with missing or noisy data. A reference for further reading on decision trees can be found in the book "Data Mining: Concepts and Techniques" by Jiawei Han, Micheline Kamber, and Jian Pei (Morgan Kaufmann, 2011).

جامعه، نمونه (حجم نمونه) و روش نمونه‌گیری

جامعه آماری این پژوهش را کلیه دانشجویان دانشگاه تهران در پنج سال گذشته (۱۳۹۵-۱۴۰۰) تشکیل می‌دهند. از میان با استفاده از روش تمام سرشماری اطلاعات دانشجویان دانشکده‌های؛ روان‌شناسی و علوم تربیتی، مدیریت، اقتصاد و مجموعه

دانشکده‌های فنی و مهندسی (۱۱ دانشکده فنی و مهندسی) در پنج سال گذشته گردآوری و تحلیل شدند. در واقع، یکی از محدودیت‌های اصلی پژوهش عدم دسترسی به مجموعه کامل داده‌های دانشجویان تحصیلات تکمیلی در تمامی دانشکده‌های دانشگاه تهران بود.

متغیرها و نحوه کدگذاری

متغیرهای مستقل در این پژوهش شامل: دانشکده، سن، مقطع تحصیلی، نوع دوره، ملیت، وضعیت تاهل، تعداد فرزندان، سال، کمک مالی، جنسیت، تصمیم به انتقال تحصیلی، تصمیم به انصراف، تصمیم به حذف ترم، تصمیم به مرخصی، تصمیم به مهمان شدن و تصمیم به تغییر رشته و متغیر وابسته: معدل بود. در جدول زیر نوع متغیرها و شیوه کدگذاری آن‌ها آمده است:

جدول ۱: نحوه ورود متغیرهای مستقل و وابسته

متغیرهای مستقل	حالت تبدیلی	برچسب گذاری	توضیحات
مقطع تحصیلی	به صورت گسسته	عدد ۰ = کارشناسی عدد ۱ = کارشناسی ارشد عدد ۲ = دکتری	*
دانشکده	به صورت گسسته	عدد ۱ = دانشکده فنی و مهندسی عدد ۲ = روان شناسی و علوم تربیتی عدد ۳ = دانشکده اقتصاد عدد ۴ = دانشکده مدیریت	*
سن	به صورت پیوسته	-	*
نوع دوره	به صورت گسسته	عدد ۱- دانشجویان بدون شهریه عدد ۰- دانشجویان شهریه پرداز	تقسیم بندی دوره‌ها بر اساس پرداخت شهریه است. در دسته ما دانشجویانی که هر ترم مبلغی را به شهریه به دانشگاه پرداخت کرده بودند به عنوان دانشجویان شهریه پرداز برچسب گذاری شدند و دانشجویانی که هر ترم هیچ مبلغی را به عنوان شهریه پرداخت نمی کردند به عنوان دانشجویان بدون شهریه برچسب گذاری شدند. در کشور ایران، به طور کلی بر اساس آزمون سراسری (ورودی دانشگاه‌ها) رتبه‌های عالی تر برای تحصیل هزینه پرداخت نمی کنند و رتبه‌های ضعیف تر مبلغی را به عنوان شهریه هر ترم پرداخت می کنند
ملیت	به صورت گسسته	عدد ۱ = دانشجویان ایرانی عدد ۰ = دانشجویان بین المللی	*
وضعیت تاهل	به صورت گسسته	عدد ۱ = مجرد عدد ۰ = متاهل	*
تعداد فرزندان	به صورت پیوسته	-	*
سال	به صورت گسسته	عدد ۱ = ۱۳۹۵ عدد ۲ = ۱۳۹۶ عدد ۳ = ۱۳۹۷ عدد ۴ = ۱۳۹۸	*

	عدد ۵ = ۱۳۹۹ عدد ۶ = ۱۴۰۰		
کمک مالی	به صورت گسسته	عدد ۱ = عدم دریافت کمک مالی عدد ۰ = دریافت کمک مالی	کمک های مالی که از طرف دولت به دانشجویان کم برخوردار اعطا می شود
جنسیت	به صورت گسسته	عدد ۱ = مردان عدد ۰ = زنان	*
تصمیم به انتقال تحصیلی، تصمیم به انصراف، تصمیم به حذف ترم، تصمیم به مرخصی، تصمیم به مهمان شدن و تصمیم به تغییر رشته	به صورت گسسته	عدد ۱ = تصمیم گرفته شده عدد ۰ = تصمیم گرفته نشده	منظور از انتقالی، شرایطی است که یک دانشجو با موافقت دانشگاه مبدا و مقصد برای یک نیمسال یا بیشتر به دانشگاه دیگر برای ادامه تحصیل منتقل می شود (این انتقال بر اساس قوانین و مقررات خاص انجام می شود). تصمیم به انصراف وقتی است که دانشجو به دلیل گوناگون تصمیم قطعی به عدم ادامه تحصیل می گیرد. تصمیم به حذف ترم، وقتی است که یک دانشجو با دلایل موجه (برای نمونه دلایل پزشکی برای بیماری) تصمیم می گیرد یک نیمسال یا بیشتر را حذف کند (حذف نیمسال سقف مشخصی دارد)، تصمیم به مرخصی، وقتی است که دانشجو تصمیم می گیرد که برای یک نیمسال یا بیشتر از دانشگاه مرخصی می گیرد بر خلاف تصمیم به حذف در اینجا نیاز به ارائه دلایل قانونی نیست و بنا به خواست خود دانشجو انجام می شود. تصمیم به مهمان شدن وقتی که یک دانشجو با موافقت دانشگاه مبدا و مقصد تصمیم می گیرد برای یک نیمسال یا بیشتر برای یک درس یا بیشتر در دانشگاه دیگر مهمان شود. بر خلاف تصمیم انتقال، این تصمیم موقت و کوتاه مدت است. تصمیم به تغییر رشته وقتی که دانشجو در حین با توجه به شرایط و قوانین تصمیم به تغییر رشته تحصیلی خود می گیرد.
معدل	به صورت پیوسته	-	*

یافته ها

۱. بخش اول یافته های توصیفی

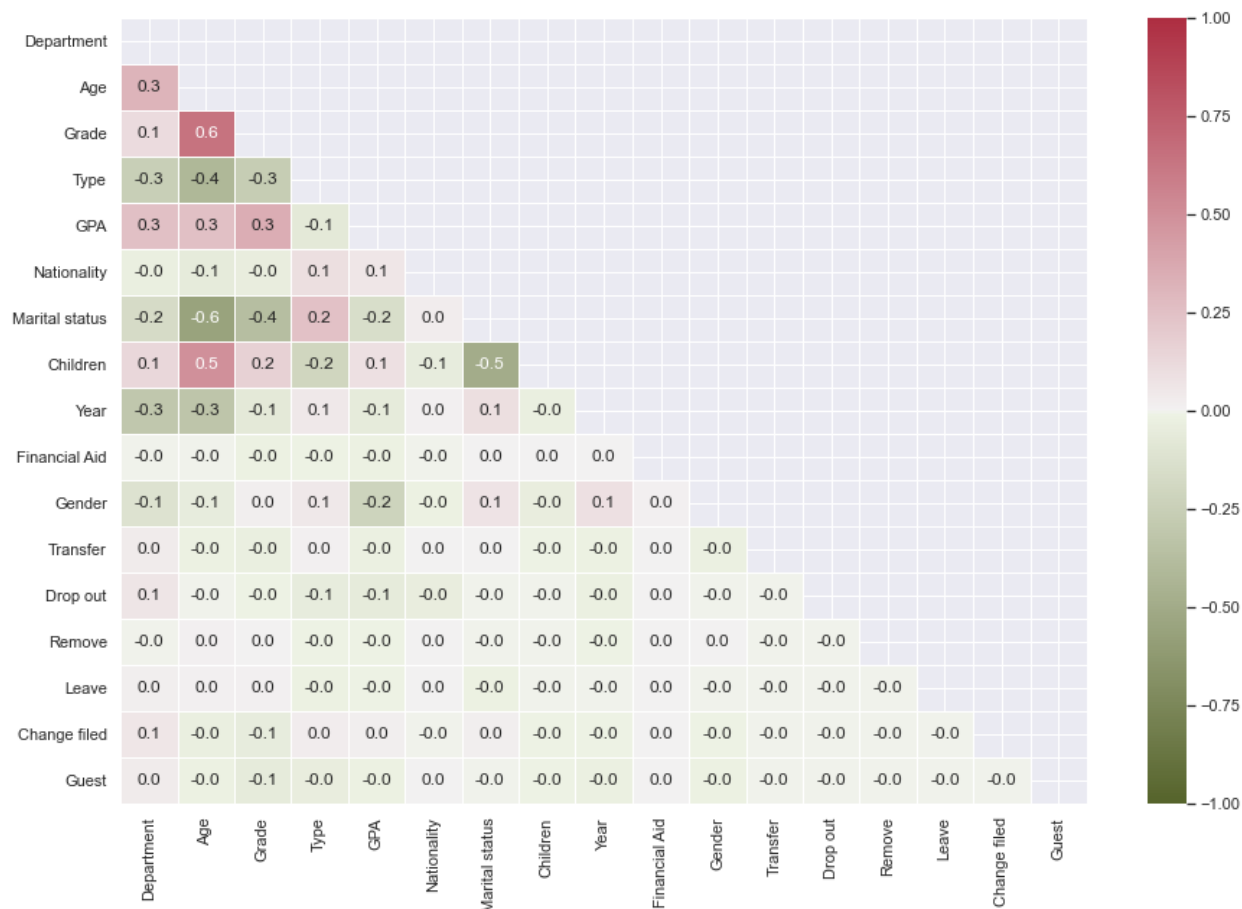
جدول: شاخص های آمار توصیفی

متغیرها	میانگین	انحراف استاندارد	کمینه	چارک اول (۲۵ درصد)	چارک دوم (۵۰ درصد)	چارک سوم (۷۵ درصد)	بیشینه
دانشکده	2.154059	1.296688	1.00	1.00	2.00	4.00	4.0
سن	26.688818	6.402831	2.00	22.00	25.00	29.00	67.0
مقطع تحصیلی	0.728208	0.661330	0.00	0.00	1.00	1.00	2.0
نوع دوره	0.759720	0.427269	0.00	1.00	1.00	1.00	1.0

معدل	16.777675	1.918553	10.04	15.67	17.17	18.25	20.0
ملیت	0.987162	0.112581	0.00	1.00	1.00	1.00	1.0
وضعیت تاهل	0.843461	0.363379	0.00	1.00	1.00	1.00	1.0
فرزندان	0.125830	0.495526	0.00	0.00	0.00	0.00	9.0
سال	4.043402	1.587015	1.00	3.00	4.00	5.00	6.0
کمک مالی	0.997155	0.053263	0.00	1.00	1.00	1.00	1.0
جنسیت	0.594865	0.490936	0.00	0.00	1.00	1.00	1.0
تصمیم به انتقال	0.001605	0.040029	0.00	0.00	0.00	0.00	1.0
تصمیم به انصراف	0.006346	0.079413	0.00	0.00	0.00	0.00	1.0
تصمیم به حذف	0.001021	0.031941	0.00	0.00	0.00	0.00	1.0
تصمیم به مرخصی	0.002261	0.047501	0.00	0.00	0.00	0.00	1.0
تصمیم به تغییر رشته	0.002626	0.051179	0.00	0.00	0.00	0.00	1.0
تصمیم به مهمان شدن	0.002553	0.050465	0.00	0.00	0.00	0.00	1.0

در جدول فوق آمار توصیفی ۱۷ متغیر مورد بررسی در پژوهش آمده است. از میان متغیرها، متغیر معدل با میانگین ۱۶,۷۷۷۶۷۵، با انحراف استاندارد ۱,۹۱۸۵۵۳، مقدار کمینه و بیشینه ۱۰,۰۴ و ۲۰,۰ می باشد.

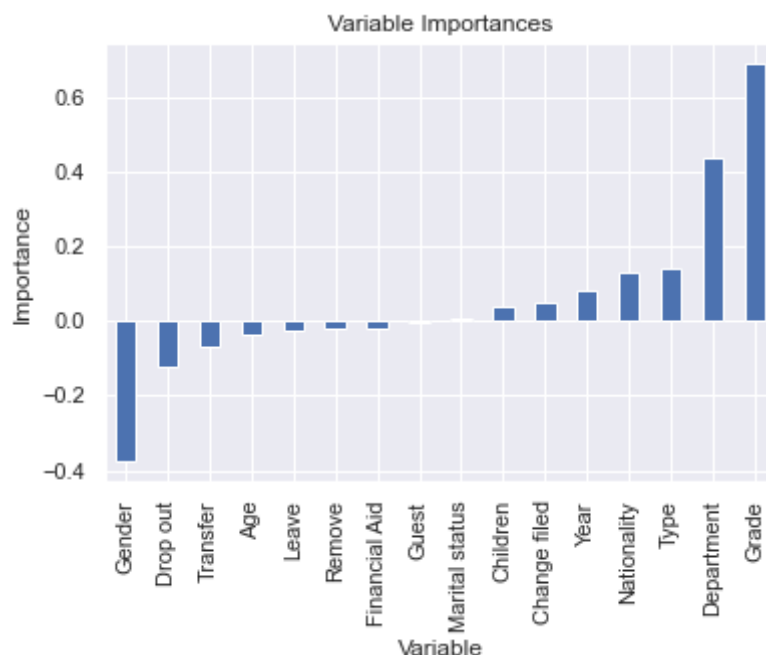
۲. بخش دوم یافته های مشترک بین مدل ها



بر اساس جدول بالا، متغیر معدل بالاترین میزان همبستگی را با متغیرهای دانشکده، سن و مقطع تحصیلی با مقدار ۳۰ درصد دارد. همچنین، کمترین میزان همبستگی با متغیرهای تصمیم به حذف، تصمیم به مرخصی و تصمیم به تغییر رشته بود.

۳. یافته های رگرسیون خطی چندگانه

در شکل زیر اهمیت متغیرهای در رگرسیون خطی چندگانه بازنمایی شده است. بر اساس این شکل، مهمترین متغیرهای در یادگیری مدل با همبستگی مثبت شامل، مقطع تحصیلی و دانشکده با مقادیر ۰٫۶۹ و ۰٫۴۳ بود. علاوه بر این، متغیر جنسیت با مقدار ۰٫۳۷ - مهمترین متغیری بود که همبستگی منفی با متغیر معدل داشت.



شکل: اهمیت متغیرها در رگرسیون خطی چندگانه

در ادامه در جدول زیر نتایج کاربست رگرسیون خطی چندگانه بیان شده است.

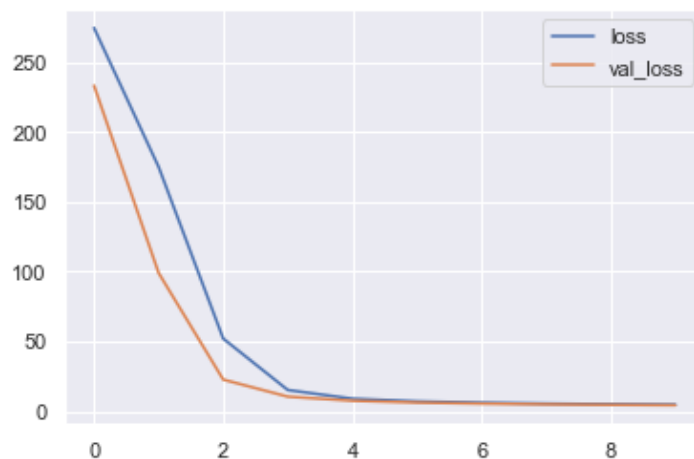
جدول : نتایج کاربست رگرسیون خطی چندگانه

مقدار	نشانگر
۱,۲۸۷۷۲۲۸۵۷۴۴۴۳۰۶۳	MAE
۲,۷۷۶۱۶۵۹۲۶۸۳۰۷۱۱	MSE
۱,۶۶۶۱۸۳۰۴۱۲۱۴۴۷۳۴	RMSE

بر اساس جدول بالا، مقصود از MAE میانگین تفاضل مقدار پیش بینی شده بین مدل و مقدار واقعی است که در این جا ۱,۲۸ بدست آمد. منظور از MSE میانگین مجذور تفاضل مقدار پیش بینی شده مدل با مقدار واقعی است که در اینجا ۲,۷۷ بدست آمد. منظور از RMSE ریشه مجذور میانگین تفاضل مقدار پیش بینی مدل با مقدار واقعی است که در اینجا ۱,۶۶ بدست آمد. مقادیر این نشانگرها هرچقدر به صفر نزدیک تر باشد نشان می دهد نشان می دهد از برازش مطلوب تری برخوردار است.

یافته های رگرسیون شبکه عصبی

در شکل زیر میزان تغییرات LOST FUNCTION آمده است. در واقع، LOST FUNCTION برای داده های تمرینی به رنگ آبی و برای داده های آزمایش به رنگ نارنجی است. روند کاهشی در تفاوت مقادیر پیش بینی شده و مقادیر واقعی بازگو کننده بالا رفتن دقت مدل شبکه عصبی در هر دوره آموزش مدل است.



شکل : نمودار LOST FUNCTION

در ادامه در جدول زیر نتایج کاربست رگرسیون شبکه عصبی بیان شده است.

جدول : نتایج کاربست رگرسیون شبکه عصبی

مقدار	نشانهگر
۱,۲۸۷۷۲۲۸۵۷۴۴۳۰۶۳	MAE
۲,۷۷۶۱۶۵۹۲۶۸۳۰۷۱۱	MSE
۱,۶۶۶۱۸۳۰۴۱۲۱۴۴۷۳۴	RMSE

بر اساس جدول بالا، مقصود از MAE میانگین تفاضل مقدار پیش بینی شده بین مدل و مقدار واقعی است که در این جا ۱,۲۸ بدست آمد. منظور از MSE میانگین مجذور تفاضل مقدار پیش بینی شده مدل با مقدار واقعی است که در اینجا ۲,۷۶ بدست آمد. منظور از RMSE ریشه مجذور میانگین تفاضل مقدار پیش بینی شده مدل با مقدار واقعی است که در اینجا ۱,۶۶ بدست آمد. مقادیر این نشانگرها هرچه قدر به صفر نزدیک تر باشد نشان می دهد نشان می دهد از برازش مطلوب تری برخوردار است.

بهینه سازی های پیرامون رگرسیون شبکه عصبی

در جدول زیر ساختار رگرسیون شبکه عصبی آمده است.

جدول: ساختار رگرسیون شبکه عصبی

Layer (type)	Output Shape	Param #
dense_285 (Dense)	(None, 128)	2176
dropout (Dropout)	(None, 128)	0
dense_286 (Dense)	(None, 128)	16512
dropout_1 (Dropout)	(None, 128)	0

dense_287 (Dense)	(None, 128)	16512
dropout_2 (Dropout)	(None, 128)	0
dense_288 (Dense)	(None, 1)	129

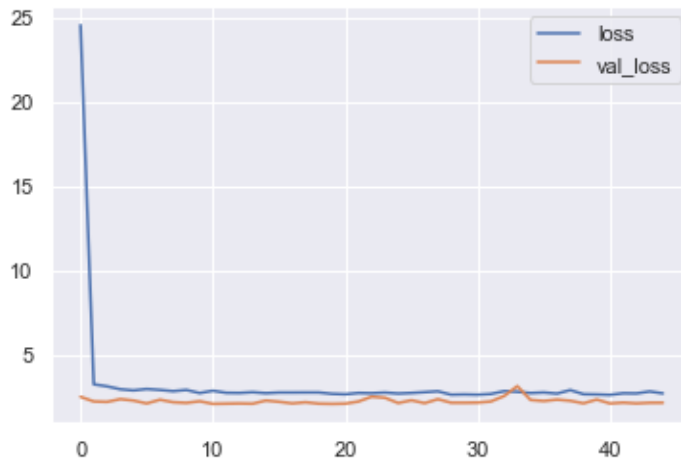
Total params: 35,329

Trainable params: 35,329

Non-trainable params: 0

بر اساس این جدول، ساختار شبکه عصبی استفاده شده روی هم رفته دارای دو لایه میانی با میزان افت ۱۰ درصد و یک لایه انتهایی باینری برای پیش بینی وجود دارد.

در شکل زیر میزان تغییرات LOST FUNCTION آمده است. در واقع، LOST FUNCTION برای داده‌های تمرینی به رنگ آبی و برای داده‌های آزمایش به رنگ نارنجی است. روند کاهشی در تفاوت مقادیر پیش بینی شده و مقادیر واقعی بازگو کننده بالا رفتن دقت مدل شبکه عصبی در هر دوره آموزش مدل است.



شکل : نمودار LOST FUNCTION

در ادامه در جدول زیر نتایج کاربست رگرسیون شبکه عصبی بهینه شده بیان شده است.

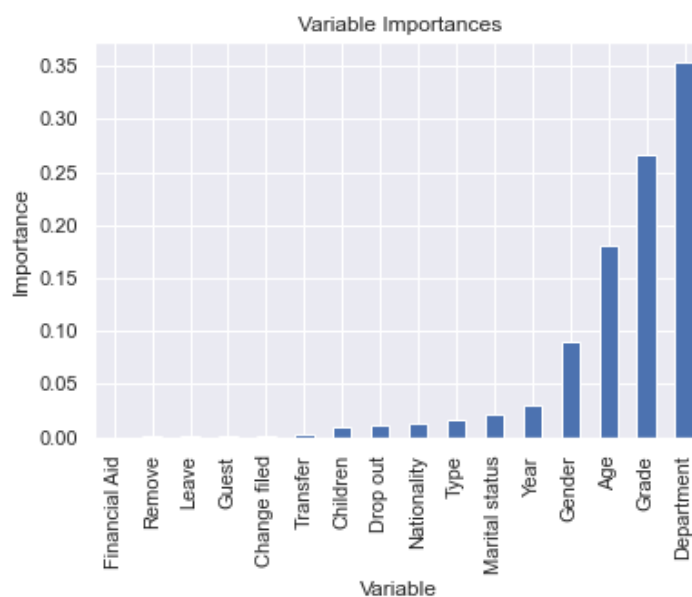
جدول : نتایج کاربست رگرسیون شبکه عصبی بهینه

مقدار	نشانگر
۱,۰۶۴۲۱۵۹۶۴۹۳۰۳۹۸۲	MAE
۲,۱۵۳۴۸۵۵۹۴۸۵۲۱۵۱۴	MSE
۱,۴۶۷۴۷۵۹۲۶۴۹۸۳۳۶۶	RMSE

بر اساس جدول بالا، مقصود از MAE میانگین تفاضل مقدار پیش بینی شده بین مدل و مقدار واقعی است که در این جا ۱,۰۶ بدست آمد. منظور از MSE میانگین مجذور تفاضل مقدار پیش بینی شده مدل با مقدار واقعی است که در اینجا ۲,۱۵ بدست آمد. منظور از RMSE ریشه مجذور میانگین تفاضل مقدار پیش بینی شده مدل با مقدار واقعی است که در اینجا ۱,۴۶ بدست آمد. مقادیر این نشانگرها هرچه قدر به صفر نزدیک تر باشد نشان می دهد نشان می دهد از برازش مطلوب تری برخوردار است.

الگوریتم جنگل تصادفی

در شکل زیر اهمیت متغیرهای در الگوریتم جنگل تصادفی بیان شده است. بر اساس این شکل، مهمترین متغیرهای در یادگیری مدل با همبستگی بالا شامل، مقطع تحصیلی و دانشکده بود.



شکل: اهمیت متغیرها در الگوریتم جنگل تصادفی

در ادامه در جدول زیر نتایج کاربست الگوریتم جنگل تصادفی شده بیان شده است.

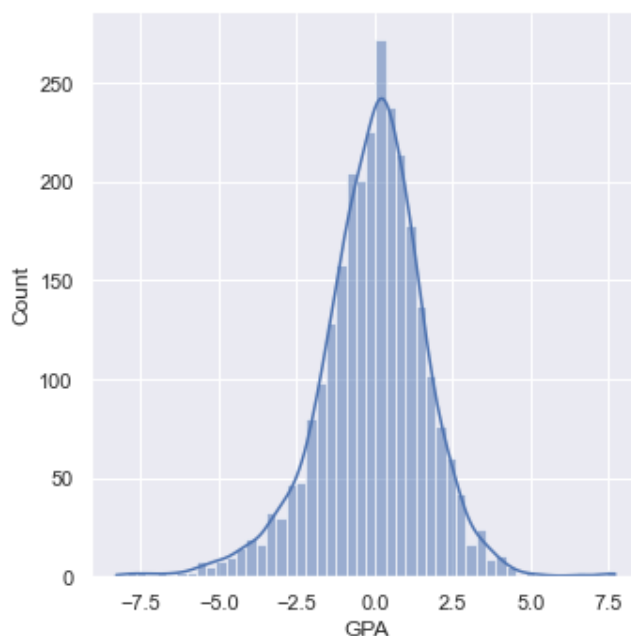
جدول : نتایج کاربست الگوریتم جنگل تصادفی

مقدار	نشانگر
۱,۲۳۳۶۸۵۸۱۹۸۳۹۰۷۵۷	MAE
۲,۵۸۱۱۳۴۷۷۷۱۲۵۳۴۵	MSE
۱,۶۰۶۵۹۱۰۴۲۲۷۷۲۰۱۶	RMSE

بر اساس جدول بالا، مقصود از MAE میانگین تفاضل مقدار پیش بینی شده بین مدل و مقدار واقعی است که در این جا ۱,۲۳ بدست آمد. منظور از MSE میانگین مجذور تفاضل مقدار پیش بینی شده مدل با مقدار واقعی است که در اینجا ۲,۵۸ بدست آمد. منظور از RMSE ریشه مجذور میانگین تفاضل مقدار پیش بینی مدل با مقدار واقعی است که در اینجا ۱,۶۰ بدست آمد. مقادیر این نشانگرها هرچقدر به صفر نزدیک تر باشد نشان می دهد نشان می دهد از برازش مطلوب تری برخوردار است.

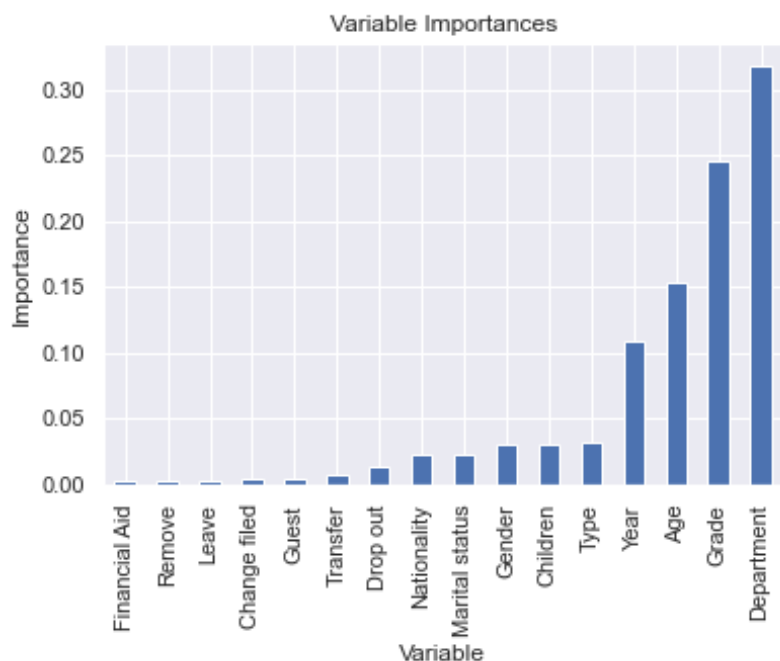
الگوریتم درخت تصمیم

نمودار زیر بیانگر میزان تفاضل مقادیر پیش بینی شده و مقادیر واقعی برای متغیر معدل است. تجمع داده ها نزدیک به عدد صفر نمایانگر میزان دقت بالای مدل می باشد.



شکل: نمودار تفاضل مقادیر پیش بینی شده و مقادیر واقعی در الگوریتم درخت تصمیم

در شکل زیر اهمیت متغیرهای در الگوریتم درخت تصمیم بیان شده است. بر اساس این شکل، مهمترین متغیرهای در یادگیری مدل با همبستگی بالا شامل، مقطع تحصیلی و دانشکده بود.



شکل: اهمیت متغیرها در الگوریتم درخت تصمیم

در ادامه در جدول زیر نتایج کاربست الگوریتم درخت تصمیم شده بیان شده است.

جدول : نتایج کاربست الگوریتم درخت تصمیم

مقدار	نشانگر
۱,۲۷۴۶۵۷۱۹۲۴۹۸۸۴۴	MAE
۲,۹۲۵۷۸۴۴۵۹۵۹۷۹۵۷	MSE
۱,۷۱۰۴۹۲۴۶۱۱۳۴۴۹۹۴	RMSE

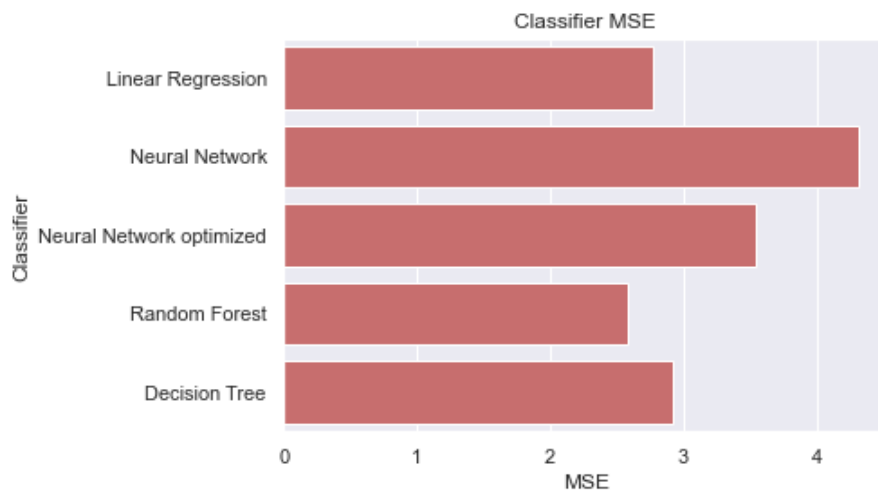
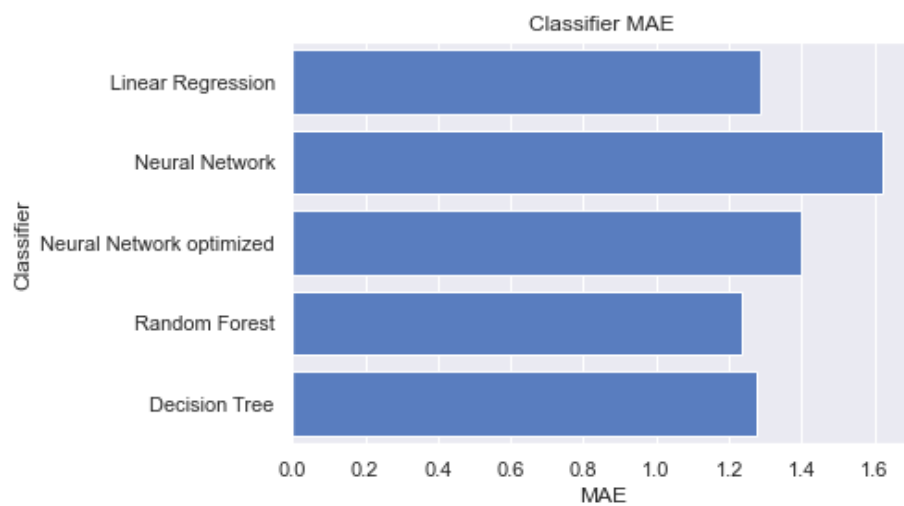
بر اساس جدول بالا، مقصود از MAE میانگین تفاضل مقدار پیش بینی شده بین مدل و مقدار واقعی است که در این جا ۱,۲۷ بدست آمد. منظور از MSE میانگین مجذور تفاضل مقدار پیش بینی شده مدل با مقدار واقعی است که در اینجا ۲,۹۲ بدست آمد. منظور از RMSE ریشه مجذور میانگین تفاضل مقدار پیش بینی مدل با مقدار واقعی است که در اینجا ۱,۷۱ بدست آمد. مقادیر این نشانگرها هرچقدر به صفر نزدیک تر باشد نشان می دهد نشان می دهد از برازش مطلوب تری برخوردار است.

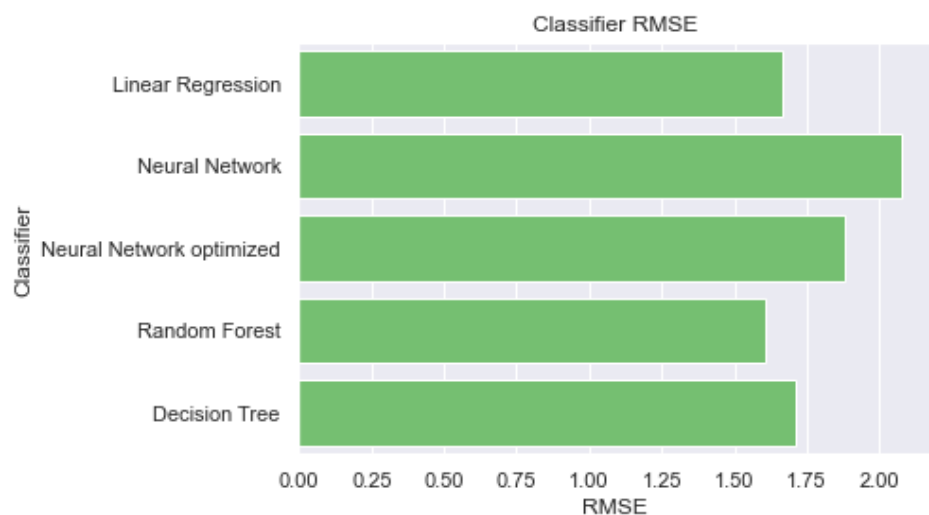
مقایسه نتایج مدل رگرسیون خطی چندگانه، رگرسیون شبکه عصبی، جنگل تصادفی و درخت تصمیم

در جدول مقایسه تطبیقی مقادیر MAE و MSE و RMSE برای تمامی الگوریتم های اجرا شده بیان شده است.

جدول: مقایسه تطبیقی الگوریتم های یادگیری نظارت شده برای پیش بینی عملکرد تحصیلی دانشجویان

الگوریتم ها	MAE	MSE	RMSE
رگرسیون خطی چندگانه	۱,۲۸	۲,۷۷	۱,۶۶
رگرسیون شبکه عصبی	۱,۲۸	۲,۷۶	۱,۶۶
رگرسیون شبکه عصبی بهینه شده	۱,۰۶	۲,۱۵	۱,۴۶
الگوریتم جنگل تصادفی	۱,۲۳	۲,۵۸	۱,۶۰
الگوریتم درخت تصمیم	۱,۲۷	۲,۹۲	۱,۷۱





بحث و نتیجه گیری

شبکه عصبی و عملکرد تحصیلی	درخت تصمیم و عملکرد تحصیلی	جنگل تصادفی و عملکرد تحصیلی
الخاصونه و هابسون ^۳ (۲۰۱۱)، دقت ۷۰,۱ درصد	کومار ^۳ و همکاران (۲۰۱۶)، دقت ۹۷,۱۷	کاوناگ و روتیکان ^۴ (۲۰۱۸) دقت ۶۶,۹٪
داس ^۵ و همکاران (۲۰۲۰)، دقت ۹۳,۷۰	بودیمان ^۶ و همکاران (۲۰۱۸)، دقت ۷۸,۵۷	الفیروز و الحاجری ^۷ (۲۰۲۰)، دقت ۷۱,۵٪
ریمال ^۸ و همکاران (۲۰۲۲)، دقت ۸۸,۵	ویونو ^۹ و همکاران (۲۰۲۰)، دقت ۹۳ درصد	نیونکیو و تانگگام ^{۱۰} (۲۰۲۰)، دقت ۴۱,۷۰
لائو ^۱ و همکاران (۲۰۱۹)، دقت ۸۴,۴	ویونو و ابیدین ^۲ (۲۰۱۹)، دقت ۹۳ درصد	پرهیزکار ^{۲۳} و همکاران (۲۰۲۳)، دقت ۷۳,۵

- | | |
|-------------------------------------|---|
| ¹ Alkhasawneh & Hobson | 2 |
| ¹ Kumar | 3 |
| ¹ Kaunang & Rotikan | 4 |
| ¹ Das | 5 |
| ¹ Budiman | 6 |
| ¹ Al-Fairouz & Al-Hagery | 7 |
| ¹ Rimal | 8 |
| ¹ Wiyono | 9 |
| ² Nuankaew & Thongkam | 0 |
| ² Lau | 1 |
| ² Wiyono & Abidin | 2 |
| ² Parhizkar | 3 |

موراتوف ^۶ و همکاران (۲۰۱۷)، دقت ۷۷ درصد	چیواپراکوبکیت ^۵ (۲۰۱۵)، دقت ۸۵,۱۸	سندووال-پالیس ^۴ و همکاران (۲۰۲۰)، دقت ۸۰ درصد
شاهین و ارول ^۹ (۲۰۲۳)، دقت ۸۰,۳۷ درصد	ریفانسیا ^۸ و همکاران (۲۰۲۰)، دقت ۵۴,۷	راچپوری ^۷ و همکاران (۲۰۱۵)، دقت ۸۹,۲۹
خان ^۳ و همکاران (۲۰۲۱)، دقت ۹۵ درصد	کولو و آدپو ^۳ (۲۰۱۵)، دقت ۶۶,۸ درصد	قویدر ^۲ و همکاران (۲۰۲۰)، دقت ۹۹,۴۲
ژانگ ^۵ و همکاران (۲۰۲۲)، دقت ۸۰ درصد	راما کریشنا ساجا ^۴ و همکاران (۲۰۲۱)، دقت ۹۵,۵	رودریگز-هرناندز ^۳ و همکاران (۲۰۲۱)، دقت ۸۲ درصد

فهرست منابع

Falát, L., & Piscová, T. (2022). Predicting GPA of University Students with Supervised Regression Machine Learning Models. *Applied Sciences*, 12(17), 8403.

Ramakrishna Sajja, V., Jhansi Lakshmi, P., Bhupal Naik, D. S., & Kalluri, H. K. (2021). Student Performance Monitoring System Using Decision Tree Classifier. In *Machine Intelligence and Soft Computing: Proceedings of ICMISC 2020* (pp. 393-407). Springer Singapore.

Rodríguez-Hernández, C. F., Musso, M., Kyndt, E., & Cascallar, E. (2021). Artificial neural networks in academic performance prediction: Systematic implementation and predictor evaluation. *Computers and Education: Artificial Intelligence*, 2, 100018.

Kumar, S. C., Chowdary, E. D., Venkatramaphanikumar, S., & Kishore, K. K. (2016, May). M5P model tree in predicting student performance: A case study. In *2016 IEEE International Conference on Recent Trends in Electronics, Information & Communication Technology (RTEICT)* (pp. 1103-1107). IEEE.

Qwaider, S. R., Abu-Naser, S. S., & Zaout, I. S. (2020). Artificial Neural Network Prediction of the Academic Warning of Students in the Faculty of Engineering and Information Technology in Al-Azhar University-Gaza.

Rachburee, N., Punlumjeak, W., Rugtanom, S., Jaithavil, D., & Pracha, M. (2015). A prediction of engineering students' performance from core engineering course using classification. In *Information Science and Applications* (pp. 649-656). Springer Berlin Heidelberg.

Kolo, D. K., & Adepoju, S. A. (2015). A decision tree approach for predicting students academic performance.

Sandoval-Palis, I., Naranjo, D., Gilar-Corbi, R., & Pozo-Rico, T. (2020). Neural network model for predicting student failure in the academic leveling course of Escuela Politecnica Nacional. *Frontiers in Psychology*, 11, 515531.

² Sandoval-Palis	4
² Cheewaparakobkit	5
² Muratov	6
² Rachburee	7
² Rifansyah	8
² Şahin & Erol	9
³ Qwaider	0
³ Kolo & Adepoju	1
³ Khan	2
³ Rodríguez-Hernánde	3
³ Ramakrishna Sajja	4
³ Zhang	5

- Lau, E. T., Sun, L., & Yang, Q. (2019). Modelling, prediction and classification of student academic performance using artificial neural networks. *SN Applied Sciences*, 1, 1-10.
- Rimal, Y., Rathore, V. S., Pageni, S., Samanta, D., Karuppiah, M., & Singh, D. (2022, August). Neural Network based Student Grade Prediction Model. In *2022 International Conference on Machine Learning, Computer Systems and Security (MLCSS)* (pp. 22-27). IEEE.
- Das, D., Shakir, A. K., Rabbani, M. S. G., Rahman, M., Shaharum, S. M., Khatun, S., ... & Arman, M. S. (2020). A comparative analysis of four classification algorithms for university students performance detection. In *InECCE2019: Proceedings of the 5th International Conference on Electrical, Control & Computer Engineering*, Kuantan, Pahang, Malaysia, 29th July 2019 (pp. 415-424). Springer Singapore.
- Rifansyah, M., Kurniawati, A., & Supratman, N. A. (2020, September). Student Activities Recommendations to Achieve First Job Waiting Time Target of Graduates in Telkom University: Decision Tree Approach. In *Proceedings of the 2021 International Conference on Engineering and Information Technology for Sustainable Industry* (pp. 1-6).
- Cheewaparakobkit, P. (2015). Predicting student academic achievement by using the decision tree and neural network techniques. *Human Behavior, Development And Society*, 12(2), 34-43.
- Alkhasawneh, R., & Hobson, R. (2011, April). Modeling student retention in science and engineering disciplines using neural networks. In *2011 IEEE Global Engineering Education Conference (EDUCON)* (pp. 660-663). IEEE.
- Wiyono, S., & Abidin, T. (2019). Comparative study of machine learning KNN, SVM, and decision tree algorithm to predict student's performance. *International Journal of Research-Granthaalayah*, 7(1), 190-196.
- Wiyono, S., Wibowo, D. S., Hidayatullah, M. F., & Dairoh, D. (2020). Comparative study of KNN, SVM and decision tree algorithm for student's performance prediction. (*IJCSAM*) *International Journal of Computing Science and Applied Mathematics*, 6(2), 50-53.
- Khan, F., Weiss, G. M., & Leeds, D. D. (2021). Predicting the Academic Performance of Undergraduate Computer Science Students Using Data Mining. In *Advances in Software Engineering, Education, and e-Learning: Proceedings from FECS'20, FCS'20, SERP'20, and EEE'20* (pp. 303-317). Springer International Publishing.
- Budiman, E., Kridalaksana, A. H., & Wati, M. (2018). Performance of decision tree C4. 5 algorithm in student academic evaluation. In *Computational Science and Technology: 4th ICCST 2017*, Kuala Lumpur, Malaysia, 29–30 November, 2017 (pp. 380-389). Springer Singapore.
- Zhang, W., Wang, Y., & Wang, S. (2022). Predicting academic performance using tree-based machine learning models: A case study of bachelor students in an engineering department in China. *Education and Information Technologies*, 1-16.
- Şahin, S., & Erol, Ç. (2022). Prediction of Secondary School Students' Academic Achievements with Machine Learning Methods and a Sample System. *Cybernetics and Systems*, 1-21.
- Muratov, E., Lewis, M., Fourches, D., Tropsha, A., & Cox, W. C. (2017). Computer-assisted decision support for student admissions based on their predicted academic performance. *American journal of pharmaceutical education*, 81(3).
- Parhizkar, A., Tejeddin, G., & Khatibi, T. (2023). Student performance prediction using datamining classification algorithms: Evaluating generalizability of models from geographical aspect. *Education and Information Technologies*, 1-19.
- Nuankaew, W., & Thongkam, J. (2020, June). Improving student academic performance prediction models using feature selection. In *2020 17th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)* (pp. 392-395). IEEE.

- Al-Fairouz, E. I., & Al-Hagery, M. A. (2020). The most efficient classifiers for the students' academic dataset. *International Journal of Advanced Computer Science and Applications*, 11(9).
- Kaunang, F. J., & Rotikan, R. (2018, October). Students' academic performance prediction using data mining. In 2018 Third International Conference on Informatics and Computing (ICIC) (pp. 1-5). IEEE.
- Singh, S. P., Malik, S., & Singh, P. (2016). Research paper factors affecting academic performance of students. *Indian Journal of Research*, 5(4), 176-178.
- Ali, N., Jusof, K., Ali, S., Mokhtar, N., & Salamat, A.S.A. (2009). THE FACTORS INFLUENCING STUDENTS' PERFORMANCE AT UNIVERSITI TEKNOLOGI MARA KEDAH, MALAYSIA. *Management Science and Engineering*, 3(4), P 81-90
- Khan, F., Weiss, G. M., & Leeds, D. D. (2021). Predicting the Academic Performance of Undergraduate Computer Science Students Using Data Mining. In *Advances in Software Engineering, Education, and e-Learning: Proceedings from FECS'20, FCS'20, SERP'20, and EEE'20* (pp. 303-317). Springer International Publishing.
- ALLAH, A. Q. G. F. (2019). Using machine learning to support students' academic decisions (Doctoral dissertation, The British University in Dubai (BUiD)).
- Putpuek, N., Rojanaprasert, N., Atcharyachanvanich, K., & Thamrongthanyawong, T. (2018, June). Comparative study of prediction models for final GPA score: a case study of Rajabhat Rajanagarindra University. In 2018 IEEE/ACIS 17th International Conference on Computer and Information Science (ICIS) (pp. 92-97). IEEE.
- Maulana, M. F., & Defriani, M. (2020). Logistic model tree and decision tree J48 algorithms for predicting the length of study period. *PIKSEL: Penelitian Ilmu Komputer Sistem Embedded and Logic*, 8(1), 39-48.
- Ahmad, F., Ismail, N. H., & Aziz, A. A. (2015). The prediction of students' academic performance using classification data mining techniques. *Applied mathematical sciences*, 9(129), 6415-6426.
- Rehman, W., Degirmen, S., & Waseem, F. (2021). Propensity for and Quality of Intellectual Capital Divulgence Across the BRICS Banking Sector: A Knowledge-Based Perspective from Emerging Economies. *Journal of the Knowledge Economy*, 1-28.
- Phale, K., Fanglin, L., Adjei Mensah, I., Omari-Sasu, A. Y., & Musah, M. (2021). Knowledge-Based Economy Capacity Building for Developing Countries: A Panel Analysis in Southern African Development Community. *Sustainability*, 13(5), 2890.
- Ngepah, N., Saba, C. S., & Mabindisa, N. G. (2021). Human capital and economic growth in South Africa: A cross-municipality panel data analysis. *South African Journal of Economic and Management Sciences*, 24(1), 11.