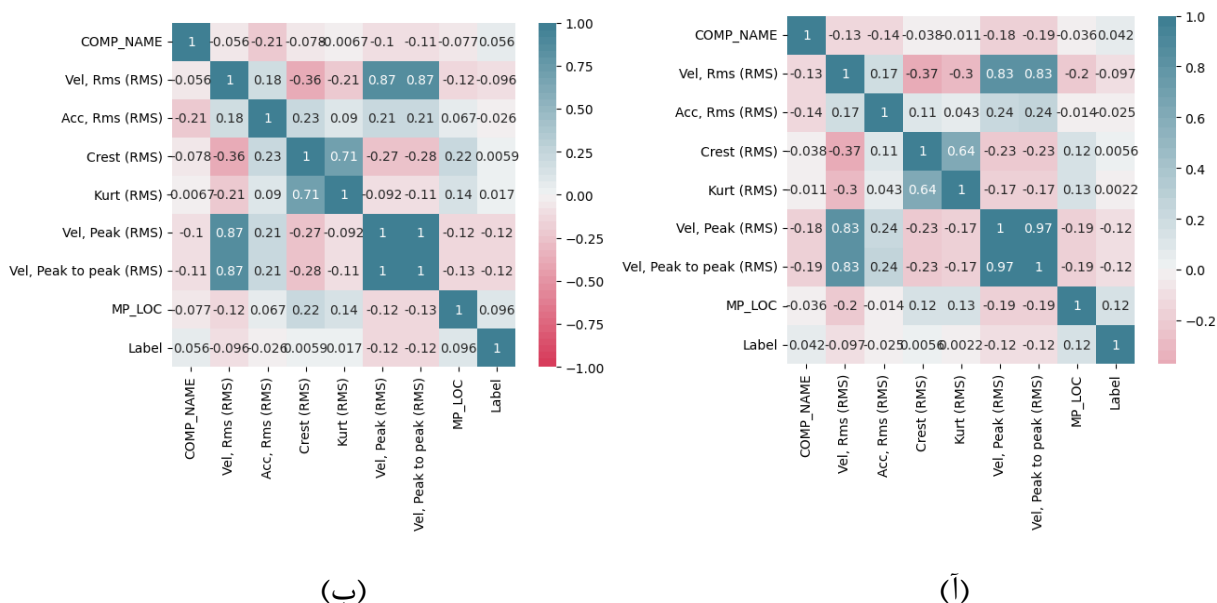


عنوان: راه حل ارائه شده برای پیشبینی خرابی بیرینگ تجهیزات الکتروفن با استفاده از داده های ارتعاشی

مرحله اول: بررسی داده

در این مرحله ابتدا داده های در اختیار قرار گرفته با استفاده از یک تابع عملکردی مورد بررسی قرار خواهد گرفت در این فایل با توجه به تعداد ۱۳۸۵ نمونه موجود در منبع داده به بررسی نمونه هایی با داده خالی و همچنین بررسی وجود و یا عدم وجود داد های تکراری، داده هایی با رابطه خطی کامل با هدف مورد بررسی قرار گرفته است. سپس به بررسی رابطه پیرسون بین ویژگی های ورودی و تابع هدف پرداخته ایم. نتیجه این بررسی در تصویر ۱ نشان داده شده است.



تصویر ۱. رابطه بین ویژگی های ورودی و هدف نهایی با استفاده از : (آ) تابع کندال^۱، (ب) تابع پیرسون^۲.

همانگونه که در تصویر ۱ نشان داده شده است، رابطه بین ویژگی ها و تابع هدف بصورت خطی در بهترین شرایط ۰.۰۹۶ و متعلق به "MP_LOC" و هدف منبع داده است. در بین روابط غیر خطی بین ویژگی هایی ورودی و مقدار تابع هدف نیز "MP_LOC" رابطه ای با میزان ۰.۱۲ بصورت غیر خطی با هدف را داراست. با توجه به بررسی های به عمل آمده ویژگی هایی نظیر سرعت، سرعت پیک و سرعت پیک به ترتیب دارای روابط ۰.۸۷٪، ۰.۱۰٪ و ۰.۱۰٪ با یکدیگر هستند. قدم بعدی در بررسی داده انتخاب ویژگی های مناسب برای پیش

^۱ Kendall

^۲ Pearson

بینی نهایی است. در این مرحله با توجه به وجود رابطه منفی بین بعضی از ویژگی ها و رابطه صد در صدی آنها با یکدیگر تصمیم گرفته شده است تا ویژگی های جدیدی بر اساس ویژگی های سرعت دورانی و شتاب دورانی ایجاد شود. ویژگی های جدید ایجاد شده شامل تقسیم شتاب بر سرعت و ضرب شتاب در سرعت است. سپس این موارد به منبع داده برای پیش بینی نهایی اضافه شده است.

در این مرحله علاوه بر اضافه کردن ویژگی های مختلف نمونه هایی که منجر به ایجاد ناهنجاری در منبع داده شده است نیز حذف شده اند. در این مرحله اگر نمونه ای میزان سرعت آن بیشتر از ۶۰۷ و یا سرعت پیک آن بیشتر از ۱۷۰۳۳ دور بر ثانیه باشد و یا سرت دورانی پیک تا پیک آن بیشتر از ۳۲۰۷۰ دور بر ثانیه باشد از منبع داده حذف خواهد شد. دیگر دلایل حذف شدن از منبع داده در داخل کد مشخص شده است.

مرحله دوم: انتخاب ویژگی

در این مرحله ابتدا تابه پراکندگی منبع داده با استفاده از پراکندگی سه کلاس ۰،۱ و ۲ مورد بررسی قرار گرفته است. نتیجه این پراکندگی در جدول ۱ نشان داده است.

جدول ۱: پراکندگی تابع هدف برای منبع داده اولیه در حالت آموزش مدل.

کلاس "۰"	کلاس "۱"	کلاس "۲"
۳۷٪	۳۸٪	۲۵٪

همانطور که در جدول ۱ نشان داده است، پراکندگی منبع داده در مقادیر هدف به صورت یکسان نبوده است. در این حالت ابتدا منبع داده با استفاده از روش نمونه برداری افزایشی^۳ تقویت شده است. قابل ذکر است که این عمل تقویتی تنها برای مقادیر داده های آموزشی انجام شده است. نتیجه این اعمال تقویتی میزان پراکندگی مساوری با ۳۳٪ برای تمامی کلاس های موجود در منبع داده است. لازم به ذکر است میزان پراکندگی تابع هدف در قسمت آموزشی به صورت ثابت باقی مانده است و مقادیر این پراکندگی در جدول ۲ ذکر شده است.

جدول ۲: پراکندگی تابع هدف برای منبع داده اولیه در حالت اعتبار سنجی مدل.

کلاس "۰"	کلاس "۱"	کلاس "۲"
۳۸٪	۳۷٪	۲۵٪

³ Border line SMOTE

در مرحله بعدی با استفاده از تابع نرمال سازی، مقادیر ورودی موجود در منبع داده در دو حالت آموزشی و اعتبار سنجی به فرآیند نرمال سازی رسیده است. سپس برای انتخاب ویژگی از روش رو به جلو انتخاب ویژگی^۴ و مدل جنگل تصادفی^۵ و ماشین بردار پشتیبان^۶ استفاده شده است. نتیجه این فرآیند آموزشی در تصویر ۲ نشان داده شده است.

index	feature_idx	cv_scores	avg_score	feature_names
1	7	[0.4254386 0.4122807 0.43612335 0.39207048 0.51101322]	0.4353852693407528	7
2	0,7	[0.56578947 0.60087719 0.59030837 0.52422907 0.66519824]	0.58928046989721	0,7
3	0,1,7	[0.55263158 0.62719298 0.64317181 0.59911894 0.69281938]	0.6209899387124198	0,1,7
4	0,1,3,7	[0.57017544 0.6754398 0.65198238 0.64757709 0.67400881]	0.6438384834052081	0,1,3,7
5	0,1,3,4,7	[0.60087719 0.65789474 0.65638767 0.67841411 0.67841411]	0.6543975577710797	0,1,3,4,7
6	0,1,3,4,5,7	[0.57894737 0.67982456 0.63436123 0.65198238 0.72687225]	0.6543975577710797	0,1,3,4,5,7
7	0,1,2,3,4,5,7	[0.59210526 0.65350877 0.66079295 0.66079295 0.74449339]	0.6623386660483808	0,1,2,3,4,5,7
8	0,1,2,3,4,5,6,7	[0.5877193 0.65789474 0.62114537 0.64317181 0.77092511]	0.6561712651673236	0,1,2,3,4,5,6,7

(آ)

index	feature_idx	cv_scores	avg_score	feature_names
1	7	[0.4254386 0.40789474 0.43612335 0.39207048 0.51101322]	0.4345080763582966	7
2	0,7	[0.54824561 0.59649123 0.6123348 0.54625551 0.64317181]	0.5892997913285416	0,7
3	0,3,7	[0.54824561 0.63596491 0.64317181 0.59911894 0.66079295]	0.6174588453512637	0,3,7
4	0,2,3,7	[0.54824561 0.65789474 0.69603524 0.6123348 0.69603524]	0.6421091274441612	0,2,3,7
5	0,2,3,4,7	[0.5877193 0.67105263 0.66079295 0.63436123 0.72246696]	0.6552786150398021	0,2,3,4,7
6	0,2,3,4,6,7	[0.60964912 0.65350877 0.64317181 0.64757709 0.74889868]	0.6605610943658706	0,2,3,4,6,7
7	0,1,2,3,4,6,7	[0.58333333 0.64473684 0.62995595 0.64757709 0.73127753]	0.6473761496251642	0,1,2,3,4,6,7
8	0,1,2,3,4,5,6,7	[0.57894737 0.6622807 0.6123348 0.66519824 0.74008811]	0.6517698431099775	0,1,2,3,4,5,6,7

(ب)

تصویر ۲. روابط بین ویژگی های ورودی و دقت بدست آمده برای پیش بینی کلاس های تابع نهایی با استفاده از: (آ) ماشین بردار پشتیبان، جنگل تصادفی.

با توجه به نتایج نشان داده شده در تصویر ۲، همانطور که مشخص است، بهترین ویژگی به منظور پیش بینی هدف ویژگی ۷ یا همان "MP_LOC" است. با توجه به تایید بدست آمده بهترین ویژگی های مورد استفاده برای پیش بینی مقادیر نهایی عبارتند از: "MP_LOC"، "COMP_NAME"، "Vel, Rms (RMS)"، "Crest(RMS)"، "Kurt (RMS)"، "Vel Peak (RMS)" و "Acc Rms (RMS)".

با توجه به نتایج بدست آمده با بررسی مدل به میزان ۵ بار بهترین دقت بدست آمده برابر با ۶۶.۲٪ است بر روی داده های آموزشی است. در نهایت مدل بر روس داده های تست دقتی برابر با ۵۷٪ است. که دقت مورد نیاز برای تست داده ها مناسب نیست.

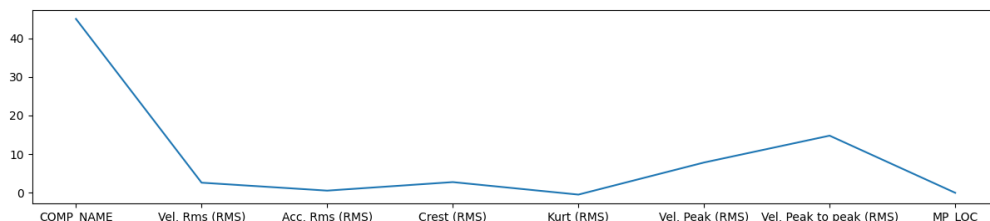
⁴ Forward Feature Selection

⁵ Random Forest

⁶ Support Vector Machine

مرحله سوم: تقویت دیتای ورودی

با توجه به داده نتایج بدست آمده در مرحله دوم، دقت حاصله با توجه به استفاده تنها از داده های ورودی بصورت سیگنال های ارتعاشی برای پیش بینی نهایی کافی نیست. در این مرحله با استفاده توابع مختلف مصور سازی اقدام به تبدیل سیگنال ورودی به تصویر کرده ایم. نمونه ای از تصاویر مورد استفاده برای آموزش در تصویر ۳ نشان داده شده است.



تصویر ۳. نمونه ای از تصویر ایجاد شده با استفاده از کشیدن تصویر خطی: (آ) تابع با هدف "۲".

همانطور که در تصویر ۳ نشان داده شده است، تصویر مورد نیاز برای ورود به سیستم تنها نمودار خطی در نظر گرفته شده برای ورود به سیستم است. استفاده از نمودار های متفاوتی نظیر نمودار نقطه ای برای ایجاد ورودی برای مدل در نظر گرفته شده است که با انجام فرآیند اعتبار سنجی نمودار های خطی به عنوان تابع ورودی منجر به رسیدن به دقتی بالاتر شده است. داده های تصاویر ایجاد شده در این مرحله به داده های سیگنال ورودی اضافه شده و بصورت همزمان برای آموزش مدل مورد نظر استفاده شده است.

مرحله سوم: آموزش مدل

مفروضات

مفروضات در نظر گرفته شده در این مرحله برای آموزش و اعتبار سنجی شبکه ها وجود پراکندگی شبیه به داده های آموزشی در اختیار قرار گرفته در داده های تست است. همچنین برای جلوگیری از کاهش دقت بین آموزش و مرحله اعتبار سنجی به نمونه هایی که متعلق به کلاس هایی با پراکندگی کمتر هستند بیشتر اهمیت داده شده است. استفاده از فرایند تقویت منبع داده با استفاده از تصاویر به پیشرفت و بهینه سازی مدل کمک خواهد کرد.

در این مرحله با توجه به اینکه ساختار مدل مورد استفاده از دو بخش تشکیل شده است، هر بخشی بصورت جداگانه مورد بررسی و تحلیل قرار خواهد گرفت.

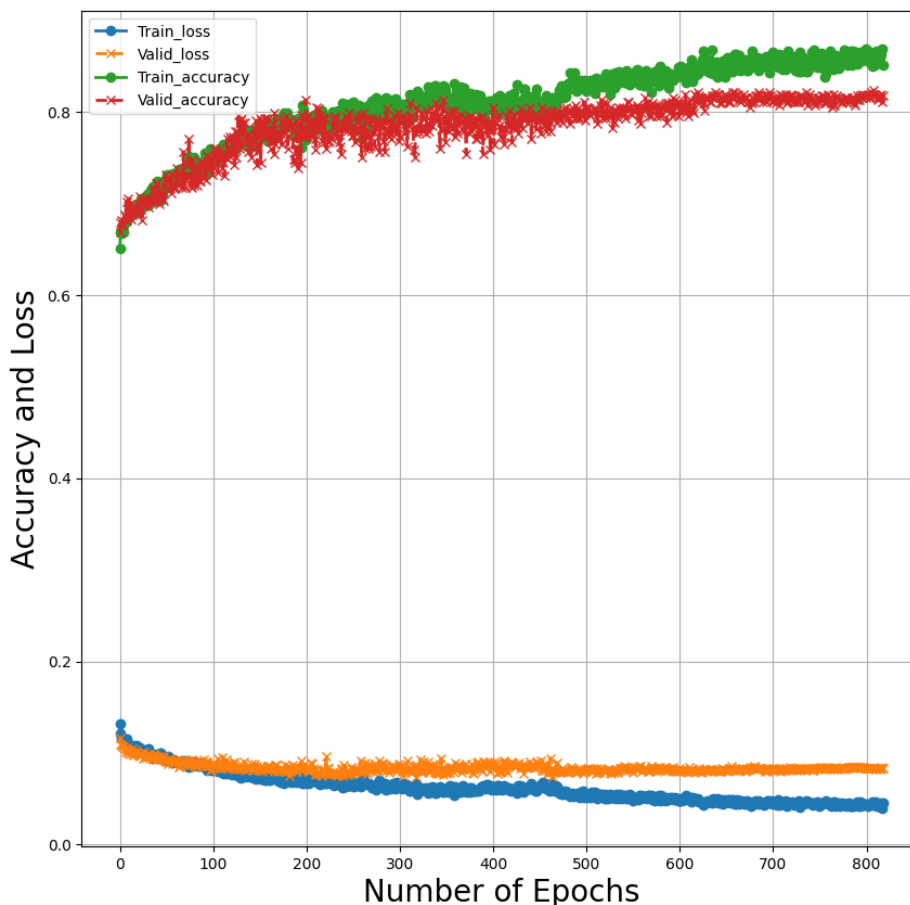
همانطور که در مرحله قبل اشاره شده است داده های ورودی در حالت سیگنال و بصورت خام در صورت وارد شدن به سیستم دارای دقت مناسبی نخواهند بود. بنابراین کلیه داده ها و داده های اضافه شده برای ورود به سیستم ابتدا تبدیل به تصویر خواهند شد و سپس با استفاده از یادگیری انتقالی به دسته بندی حالت های مختلف گفته شده خواهند پرداخت. این مدل از ساختار مدل های معروفی نظیر VGG, Resnet, Xception net و... به منظور استخراج ویژگی استفاده خواهد کرد. در نهایت برای خروجی استخراج شده در این مرحله با خروجی مرحله پیش با استفاده از لایه زنجیر وار ترکیب شده.

در بخش پیش بینی نهایی و دسته بندی ساز از دو لایه عمیق به منظور لایه پنهان و یک لایه نهایی به عنوان دسته بندی ساز نهایی استفاده خواهد شد. تابع فعال ساز در قسمت نهایی تابع softmax است.

با توجه به نکات بیان شده در این مراحل چندین احتمال متفاوت برای انتخاب فرایارامتر های موجود در شبکه وجود دارد. به منظور انتخاب هر کدام از این فرایارامتر ها از شبکه عصبی ژنتیک استفاده شده است. شبکه عصبی ژنتیک مورد استفاده در این بخش از تابع هزینه برابر با جمع مقادیر دقت با عکس خطا استفاده خواهد کرد. در نهایت به منظور آموزش مدل از بهینه ساز Adamw و تابع خطایی برابر با Focal Cross Entropy استفاده شده است.

مرحله چهارم: نتایج

در این مرحله به بررسی نتایج با استفاده از ساختار ارائه شده خواهیم پرداخت با توجه با درخواست سیستم به منظور دریافت بهترین نتیجه بر اساس جواب مثبت و جواب منفی درست، این موارد در نتیجه ارسالی در نظر گرفته شده است. نتایج بدست آمده در شکل ۴ نشان داده شده است.



شکل ۴. نتایج دقت و خطا در طول فرآیند آموزشی برای ۶۰۰ مرحله.

در این شرایط بهترین بهترین دقت برای داده تست برابر با ۸۲.۱۱٪ برای داده تست است. میزان جواب درست مثبت برابر است با ۱۵۳ و میزان جواب منفی درست برابر است با ۴۰۰ عدد. دیگر نتایج قابل ارائه در جدول ۳ نشان داده شده است. با توجه به نتایج نشان داده شده. کمترین میزان بیان شده برای جواب نادرست مثبت در حین یادگیری برابر است با ۵۵ و کمترین مفادیر ذکر شده برای جواب نادرست منفی برابر است با ۵۳.

جدول ۳: عملکرد مدل بر اساس داده های تست.

AUC	F1-score	Recall	Precision	Accuracy	Loss
۰.۸۷۴۸	۷۶.۷۳٪	۷۱.۹۵	۸۲.۲۰٪	۸۲.۳۵٪	۰.۰۸۳۲

جدول ۴: عملکرد مدل بر اساس داده های تست.

TP	TN	FP	FN
----	----	----	----

۱۵۳	۴۰۰	۵۵	۵۳
-----	-----	----	----

مرحله پنجم: مقایسه با نتایج مشابه

در این مرحله به بررسی عملکرد مدل های مشابه با استفاده از تکنیکی مشابه برای پیشبینی نهایی پرداخته ایم. نتایج استفاده از مدل های مشابه یادگیری ماشین در شکل ۵ نشان داده شده است.

Model	Accuracy	Balanced Accuracy	ROC AUC	F1 Score
LGBMClassifier	0.63	0.61	None	0.63
BaggingClassifier	0.62	0.61	None	0.62
DecisionTreeClassifier	0.62	0.60	None	0.62
ExtraTreesClassifier	0.61	0.59	None	0.61
XGBClassifier	0.61	0.59	None	0.61
RandomForestClassifier	0.58	0.57	None	0.58
LabelPropagation	0.54	0.53	None	0.54
LabelSpreading	0.54	0.53	None	0.54
ExtraTreeClassifier	0.53	0.52	None	0.53
AdaBoostClassifier	0.54	0.52	None	0.54
KNeighborsClassifier	0.49	0.47	None	0.49
SVC	0.49	0.46	None	0.48
NuSVC	0.43	0.42	None	0.43
QuadraticDiscriminantAnalysis	0.41	0.38	None	0.37
LogisticRegression	0.41	0.37	None	0.34
LinearDiscriminantAnalysis	0.40	0.37	None	0.34
PassiveAggressiveClassifier	0.38	0.37	None	0.37
NearestCentroid	0.34	0.37	None	0.31
CalibratedClassifierCV	0.40	0.36	None	0.34
RidgeClassifierCV	0.40	0.36	None	0.34
GaussianNB	0.40	0.36	None	0.33
LinearSVC	0.40	0.36	None	0.33
RidgeClassifier	0.40	0.36	None	0.33
BernoulliNB	0.40	0.36	None	0.34
Perceptron	0.38	0.35	None	0.30
DummyClassifier	0.35	0.33	None	0.18
SGDClassifier	0.28	0.30	None	0.26

شکل ۵. نتایج بدست آمده با انجام عملیات آموزش و اعتبار سنجی در مدل های مشابه.

به منظور اطمینان یافتن از اینکه بهترین مدل در این حالت انتخاب شده است، بهترین مدل با بهترین قابلیت در این مرحله با نام LGBM Classifier انتخاب شده و فرارامتر های این مدل برای رسیدن به دقتی مناسب تر دوباره تنظیم شده است. نتایج حاصل از این تنظیم در شکل ۶ نشان داده شده است.

```

1 predictions = clf.predict(X_test)
2 print(classification_report(y_test, predictions))

```

[LightGBM] [Warning] feature_fraction is set=0.9, colsample_bytree=1.0 will be ignored. Current value: feature_fraction=0.9

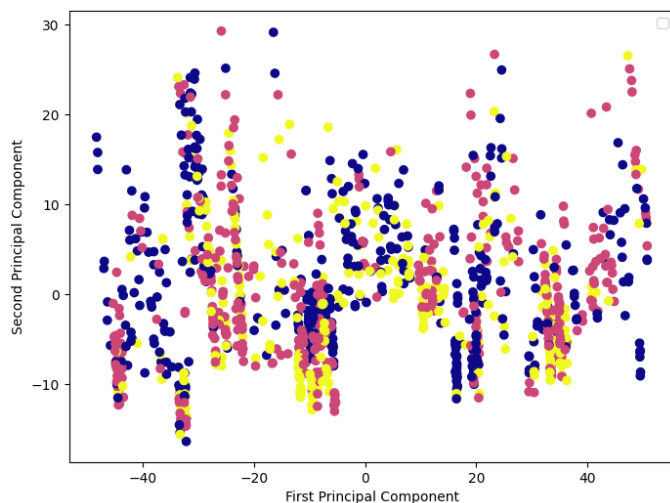
	precision	recall	f1-score	support
0	0.67	0.65	0.66	121
1	0.67	0.69	0.68	110
2	0.64	0.64	0.64	96
accuracy			0.66	327
macro avg	0.66	0.66	0.66	327
weighted avg	0.66	0.66	0.66	327

شکل ۶. نتایج بدست آمده با آموزش بهترین مدل یادگیری ماشین.

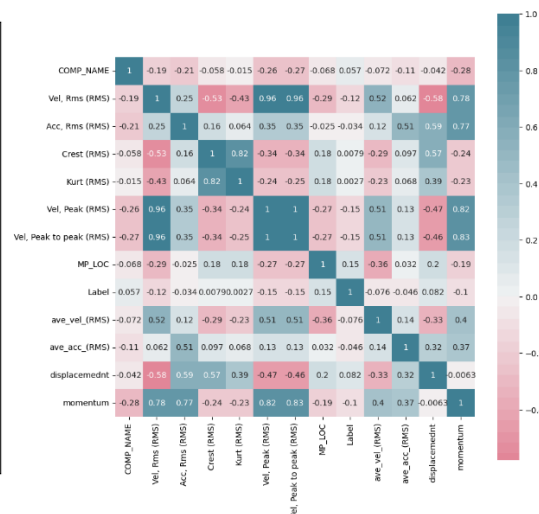
همانگونه که نشان داده شده است مدل مورد نظر دارای اختلافی ۱۰٪ با مدل ارائه شده دارد. این امر نشان دهنده قدرت مدل ارائه شده برای افزایش میزان جواب های مثبت در درست با مدل های مشابه در حین افزایش میزان پیشبینی درست منفی است.

مرحله ششم: افزایش دقت

راه حل ارائه شده در این مرحله جمع آوری هر چه بیشتر داده در مناطق مختلف است. یکی از مشکلات مشاهده شده ر نوع جمع آوری منبع داده استفاده از ویژگی هایی هست که رابطه خطی و غیر خطی نامناسبی با تابع هدف دارند. یکی از راهکارهایی برای پیشبینی بهتر و سریعتر در صنایع است. یکی دیگر از معضلات ارائه شده در دیتایی ارائه شده نزدیک و در هم تنیده بودن داده هایی است که به پیش بینی نهایی کمک میکنند. به عنوان مثال ویژگی سرت دورانی پیک و سرعت دورانی پیک تا پیک با تابع هدف رابطه خطی و غیر خطی منفی دارند و حضورشان در منبع داده هم به داده های نامناسب موجود کمک میکند و هم دقت مورد نظر را کاهش خواهد داد. در صورتیکی که ویژگی های ایجاد شده توسط گروه دارای رابطه مناسب و مثبت تری برای پیش بینی نهایی دارند. نمونه از ویژگی های ایجاد شده در شکل ۷ نشان داده شده است.



(ب)



(آ)

شکل ۷. (آ) رابطه بین ویژگی های تولید شده و تابع هدف، (ب) پراکندگی نمونه های موجود در منبع داده به تفکیک کلاس های صفر و یک و دو.

مرحله هفتم: کاربرد سازی راه حل

ارائه مدل مورد نظر در صنعت نیازمند روابط و ابزار مناسب است. استفاده از یک سیستم پیش بینی نیازمند دریافت دقیق اطلاعات به صورت روزانه و آموزش همزمان مدل ارائه شده است. استفاده از یک منبع داده مناسب و داشبورد تحلیلی برای اطلاع از میزان و نوع خرابی ها برای بررسی نحوه عملکرد مورد بررسی قرار خواهد بود. استفاده از سیستم مانیتورینگ همزمان و بررسی یک برنامه جامع برای تعمیر و نگهداری قطعات معیوب یکی از راه های ارائه شده است. استفاده از ساختار های هوش مصنوعی نظیر فرآیند تصمیم گیری هوش مند منجر به برنامه سازی منظم برای پیشبینی عملکرد سیستم است.