یافته علمی کوتاه:
پیش‌بینی صید کفال ماهیان در آبهای ایرانی دریای خزر با الگوریتم درخت تصمیم

طاهره عشورنژاد*[۱]، محمد ابراهیم شیری[۲]، مهناز ربانی‌ها[۱]

*tahereh_a2008@yahoo.com

۱- موسسه تحقیقات علوم شیلاتی کشور، سازمان تحقیقات، آموزش و ترویج کشاورزی، تهران، ایران
۲- دانشگاه صنعتی امیرکبیر، تهران، ایران

تاریخ دریافت: شهریور ۱۳۹۸ تاریخ پذیرش: اسفند ۱۳۹۸

لغات کلیدی: کفال ماهیان، صید، دریای خزر، پارامترهای زیست محیطی، الگوریتم درخت تصمیم

مقدمه

کفال ماهیان از گونه‌های مهم و اقتصادی در دریای خزر بوده و در ۱۰ سال اخیر حدود ۱۸ درصد از صید ماهیان استخوانی را در سواحل ایرانی دریای خزر بخود اختصاص داده‌اند (فضلی، ۱۳۹۷). با توجه به اهمیت کفال ماهیان در سبد صید صیادان، ماهیان استخوانی که پس از ماهی سفید از لحاظ مقدار صید و ارزش اقتصادی دارای جایگاه دوم می‌باشند، هر ساله میزان ذخایر این ماهیان در سواحل ایرانی دریای خزر برآورد می‌گردد. صید کفال ماهیان در شوروی سابق و ایران بترتیب از سال‌های ۱۹۳۷ و ۱۹۴۲ آغاز شد (Ghadirnejad, 1996؛ رضوی صیاد، ۱۳۶۹). مهم‌ترین عامل اثرگذار بر جمعیت ماهیان استخوانی بخصوص کفال ماهیان، تغییرات سطح آب دریا و درجه حرارت می‌باشد (Ivanov, 2000). تراکم جمعیت این ماهیان در ماه‌های گرم سال بشدت گسترش یافته و از تراکم بسیار بالایی برخوردار می‌شود ولی در ماه‌های سرد سال از تراکم آن کاسته می‌گردد. در ایران در دریای خزر، صید کفال ماهیان در رده دوم قرار دارد و بیش از ۱۳ درصد صید ماهیان استخوانی را بخود اختصاص می‌دهد (دریانبرد، ۱۳۸۸). صید کفال ماهیان توسط شوروی سابق از سال ۱۹۳۷ به مقدار ۶۰ تن آغاز شد و میزان صید آنها بجز سال ۱۹۵۶، همیشه کمتر از هزار تن بوده است (Ghadirnejad, 1996). صید کفال ماهیان در ایران از سال ۱۳۲۱ (۱۹۴۲ میلادی) به مقدار ۵۷ تن آغاز شد (رضوی صیاد، ۱۳۶۹). میانگین صید سالانه کفال ماهیان طی ۲۲ سال اخیر برابر ۱۰۵۱/۴± ۳۹۳۸/۳ تن با ضریب تغییرات ۲۶/۷ درصد بوده است.

Hongchun و همکاران (۲۰۱۱) سیستم پیش‌بینی شیلات بر پایه شبکه عصبی را مطرح کردند که نتایج دقیقی بدست آوردند. Rabbaniha و Gerami (۲۰۱۸) پیش بینی ارزیابی ذخایر شیلات ماهی کیلکای آنچوی در دریای خزر با استفاده از یک رویکرد سری زمانی بررسی نمودند. نتایج نشان داد که جمعیت کیلکا ماهیان در آینده بتدریج افزایش می‌یابد. خداپرست شیرازی و صادقی (۱۳۹۵) پیش بینی تولید آبزیان دریایی در ایران را با استفاده از شبکه عصبی مصنوعی ارائه دادند که نتایج خوبی بدست آوردند. امیری (۱۳۹۷) مدل سازی عوامل محیطی تاثیرگذار بر میزان صید در واحد تلاش کیلکا ماهیان را در سواحل ایرانی دریای خزر مورد بررسی قرار داد که از مدل شبکه عصبی استفاده شد و نشان داد که تراکم کلروفیل، دمای سطح آب، فصل صید و طول جغرافیایی بترتیب بیشترین تاثیر را بر میزان صید کیلکا

[Downloaded from isfj.ir on 2024-07-17]
[DOR: 20.1001.1.10261354.1399.29.5.4.0]

ماهیان جنوب دریای خزر دارد. فضلی و غنی نژاد (۱۳۸۳) به بررسی صید و برخی جنبه‌های زیست‌شناختی کفال ماهیان دریای خزر پرداختند که توانستند آمار صید کفال ماهیان را طی سال‌های ۸۱-۱۳۸۰ برآورد نمایند. فضلی و دریانبرد (۱۳۹۶) تغییرات مکانی صید و گروه بندی شرکت‌های تعاونی پره ساحلی را در سواحل ایرانی دریای خزر مورد بررسی قرار دادند که ماهی سفید ۶۶/۲۷ درصد از کل صید را بخود اختصاص داد.

با توجه به اهمیت صید و صیادی نمونه مورد نظر و تشخیص عوامل محیطی موثر بر نوسانات میزان زی‌توده و میزان صید سالانه این گونه و مشخص کردن علل تغییرات میزان زی‌توده ذخایر و پراکنش آنها و نهایتاً مدیریت بهینه و پایدار برداشت از ذخایر این گونه و پیش‌بینی صید کفال ماهیان در سال‌های آتی، تحقیق حاضر با بکارگیری نرم‌افزار Clementine18 که به نام Ibm spss modeler معروف است و از نرم افزارهای پیشرو در علم داده‌کاوی می‌باشد، انجام گردید. این نرم افزار با قابلیت جذب کاربر بوده و از روش آنالیزی الگوریتم درخت تصمیم C&R tree جهت انجام مقایسه و تعیین بیشترین دقت پیش بینی استفاده گردید (مفاهیم کلمنتاین در https://dadehkavy.com).

روش کار

داده‌های این تحقیق عمدتاً از روش کتابخانه‌ای جمع‌آوری شده است و سایر روش‌های جمع آوری بترتیب شامل مراحل ذیل بود:

- استخراج درجه حرارت آب و کلروفیل (https://neo.sci.gsfc.nasa.gov/analysis/configure.php)
- داده‌های ۱۰ ساله صید (عبدالملکی و همکاران، ۱۳۸۳؛ فضلی، ۱۳۹۷) و اطلاعات محیطی (pH، شفافیت، شوری، اکسیژن، موجودات کفزی (نصراله زاده ساروی، ۱۳۹۷) جمع‌آوری شدند.
- داده‌های مکانی حاصل از نمونه‌برداری ناحیه جنوب شرقی دریای خزر در استان مازندران و مربوط به یک صیدگاه با عرض جغرافیایی ۳۶/۵ و طول جغرافیایی ۵۳/۲ می‌باشد (شکل ۱).

شکل۱: نقشه موقعیت ایستگاه نمونه برداری مشخص شده در دریای خزر

Figure 1: Location map of sampling station specified in Caspian Sea.

[Downloaded from isfj.ir on 2024-07-17]

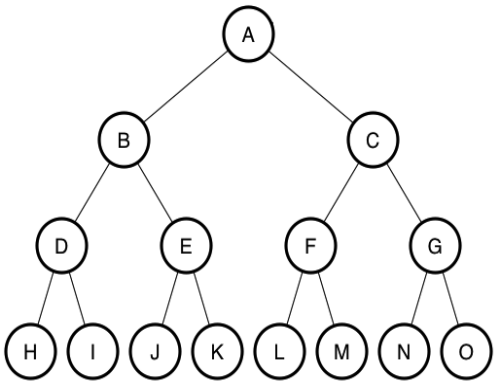
[DOR: 20.1001.1.10261354.1399.29.5.4.0]

اولین قسمت داده‌کاوی انتخاب داده و جمع‌آوری آن می‌باشد که جزء مراحل وقت‌گیر داده‌کاوی بشمار می‌آید. از مراکز مختلف داده‌ها جمع آوری گردید. سپس عملیات پردازش و پاکسازی داده‌ها انجام گرفت. داده‌ها دو دسته شامل متغیر وابسته (هدف) و متغیر مستقل (پیش بینی کننده) می‌باشند. از آنجایی که برخی داده‌ها ممکن است هیچ تأثیری در متغیر هدف نداشته باشد، حذف می‌شوند. در قدم بعدی داده‌ها تبدیل شدند. یکسری داده‌ها در بازه زمانی ماهانه و یکسری داده‌ها در بازه زمانی سالانه قرار گرفتند. برای اینکه عملیات داده‌کاوی انجام شود، بایستی یکسان‌سازی به صورت ماهانه یا سالانه انجام شود که به اصطلاح "تبدیل داده" نامیده می‌شود. یا اینکه داده‌های اعشاری تبدیل به اعداد صحیح شوند. بعد از انجام این مراحل، به فاز داده‌کاوی وارد می‌شوید. در فاز داده‌کاوی با استفاده از مدل‌های مختلف می‌توان عملیات مختلف انجام داد و در نهایت ارزیابی را از مدل خروجی بیان نمود.

درخت تصمیم C&R[1]

داده‌کاوی یک اصطلاح عمومی است که تعدادی از تکنیک‌های مورد استفاده برای شناسایی قطعات اطلاعات یا تصمیم‌گیری در داده‌ها را توصیف می‌کند (Han *et al.*, 2006). در این روش که موجب تشکیل یک درخت تصمیم با تقسیمات دوتایی می‌گردد، توسط Breiman و همکاران (۱۹۸۴) به طور کامل معرفی شد. CRT یک درخت تصمیم بشمار می‌آید که به ما این اجازه را می‌دهد که ملاحظات آینده را پیشگویی و دسته بندی نمود. این روش بر پایه قسمت بندی و جزء بندی بازگشتی است تا رکوردهای آموزشی را به داخل قطعه‌ها و اجزاء بخش بندی با کم کردن و هرس کردن بخش‌های ناخالص و غیر مهم در هر مرحله، دسته بندی و تقسیم بندی شود (Roger and Lewis, 2000) (شکل ۲).

1. Classification and regression trees



شکل۲: درخت تصمیم C&R tree
Figure 2: C&R tree decision tree.

الگوریتم پایه درخت تصمیم C&R tree (یک الگوریتم حریصانه)

- درخت به صورت بازگشتی با روش تقسیم و حل، از بالا به پایین ساخته می‌شود.
- در آغاز، همه نمونه های آموزشی در ریشه هستند .
- ویژگی‌ها چند مقداری هستند (اگر مقادیر پیوسته باشد، گسسته می‌شوند).
- داده‌ها به صورت بازگشتی و بر اساس ویژگی‌های انتخاب شده دسته‌بندی می‌شوند.
- ویژگی‌های آزمون بر اساس معیارهای اکتشافی یا آماری انتخاب می‌شوند (Han *et al.*, 2006).

نتایج

در این مقاله داده‌های محیطی و داده‌های ماهانه صید طی سال‌های ۹۰-۱۳۸۱ و داده‌های سالانه ۹۶-۱۳۸۳ مورد آنالیز قرار گرفت. برای افزایش دقت، داده‌های۱۰ سال بر حسب فصل صید و در نقطه صید مورد بررسی قرار گرفت و نتایج حاصله مطابق شکل ۳ و جدول۱ ارائه گردیده است.

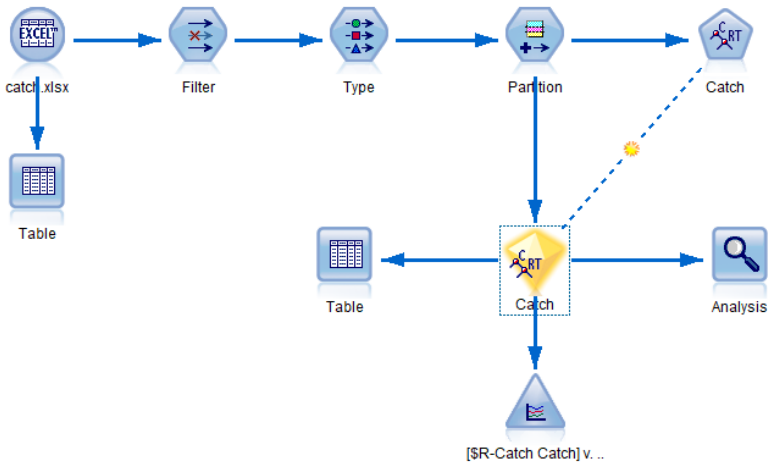
فرآیند دسته بندی در دو فاز انجام گرفت:

۱. فاز آموزش (یادگیری): در این مرحله مدل ساخته شد.
۲. فاز آزمایش (تست): در این مرحله از مدل ساخته شده برای پیش بینی متغیرها استفاده شد.

[Downloaded from isfj.ir on 2024-07-17]
[DOR: 20.1001.1.10261354.1399.29.5.4.0]

پس از آماده سازی آمارها جهت اجرای مدل درخت تصمیم، داده‌ها به دو دسته: داده‌های آموزشی و داده‌های آزمایشی تقسیم شدند. در این تحقیق ۸۰٪ از کل داده ها به آموزش مدل تعلق گرفت و ۲۰٪ باقیمانده به عنوان داده‌های آزمون به مدل معرفی گردید. انتخاب داده‌های آموزش و آزمون به صورت سیستماتیک و توسط کاربر انجام شد و بدیهی است که داده‌های آزمون در مرحله آموزش مورد استفاده قرار نگرفتند.

بعد از انجام مراحل کار و ورود داده ها در نرم افزار و انجام پیش پردازش و حذف داده‌های نویز، وقتی برنامه اجرا گردید، مدل شکل ۳ بدست آمد.



شکل۳: مدل درخت تصمیم
Figure 3: Decision Tree Model.

جدول۱: دقت الگوریتم در قسمت آموزش و آزمایش
Table 1: Algorithm accuracy in training and testing.

مدل	آموزش	آزمایش
C&R tree درخت تصمیم	٪۱۰۰	۹۸/٪۴۲

ارزیابی کارایی مدل

باید توجه داشت که مقدار واقعی هیچ کمیتی به طور دقیق مشخص نمی‌باشد. منظور از مقدار واقعی یک کمیت، مقداری است که از دقیق‌ترین آزمایش انجام شده برای اندازه‌گیری آن کمیت تاکنون گزارش شده است یا مقدار تئوری‌ای که برای یک کمیت در نظر گرفته می‌شود.

خطای نسبی

$$\text{درصد خطای نسبی } RE = \frac{|\text{نتیجه آزمایش} - \text{مقدار واقعی}|}{\text{مقدار واقعی}} \times ۱۰۰ \quad (۱)$$

[Downloaded from isfj.ir on 2024-07-17]
[DOR: 20.1001.1.10261354.1399.29.5.4.0]

میانگین خطای مطلق

$$MAE=\frac{1}{n}\sum_{i=1}^{n}|\text{نتیجه آزمایش} - \text{مقدار واقعی}| \qquad (۲)$$

میانگین مربع خطا

$$MSE=\frac{1}{n}\sum_{i=1}^{n}(\text{نتیجه آزمایش} - \text{مقدار واقعی})^2 \qquad (۳)$$

تحلیل خروجی خطا

همانطوریکه مشاهده می‌شود، خطای مدل درخت تصمیم C&R tree بر اساس جداول ۲ و ۳ در هر دو حالت ماهانه و سالانه بسیار کم می‌باشد .

جدول۲ : محاسبه خطای مدل ها بر حسب سال صید

Table 2: Calculation of models error by year of catch.

مدل	خطای نسبی	میانگین خطای مطلق	جذر میانگین مربع خطا
درخت تصمیم	۰/۱۷۹	۱/۸۱۳	۴/۷۹۶

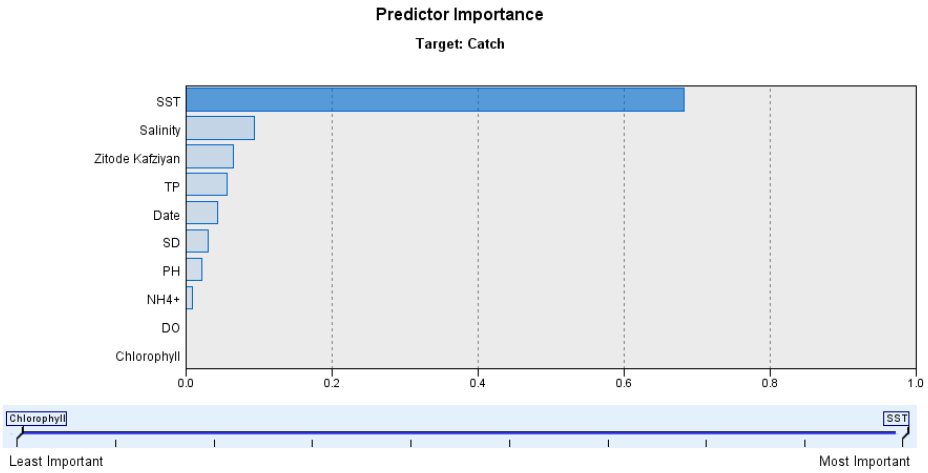
جدول۳: محاسبه خطای مدل ها برحسب ماه صید

Table 3: Calculate models error by month of catch.

مدل	خطای نسبی	میانگین خطای مطلق	جذر میانگین مربع خطا
درخت تصمیم	۰/۵۸	۱/۰۷	۳/۰۸

عوامل موثر بر صید

عوامل محیطی که بترتیب بیشترین تاثیر را بر صید کفال ماهیان دریای خزر با مدل درخت تصمیم C&R tree دارند، مطابق شکل‌های ۴ الی ۶ شامل دمای سطح آب، شوری می‌باشد.

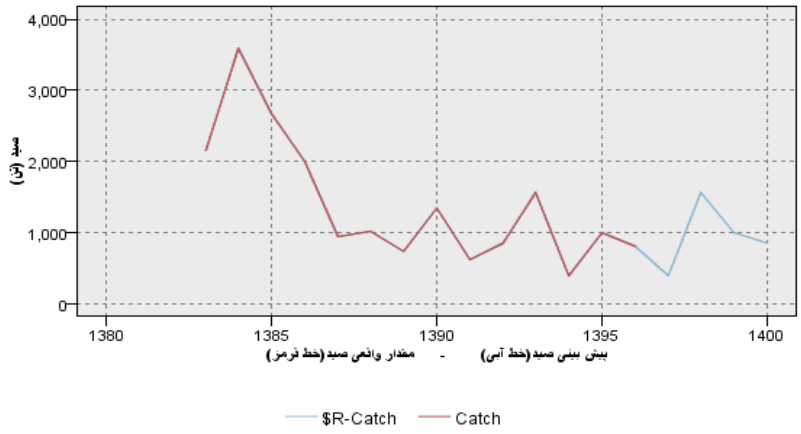


شکل۴: عوامل موثر بر صید کفال ماهیان دریای خزر

Figure 4: Factors Affecting the Caspian Sea Mullet Hunting.

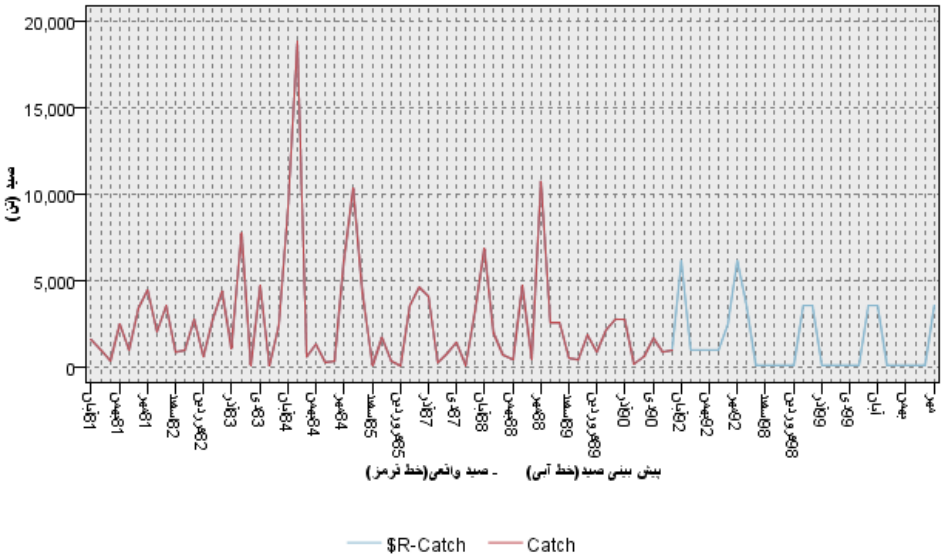
[Downloaded from isfj.ir on 2024-07-17]

[DOR: 20.1001.1.10261354.1399.29.5.4.0]



شکل۵: پیش بینی صید سالانه برای ماهی کفال

Figure 5: Prediction of annual catch for mullet.



شکل ۶: پیش بینی صید ماهانه درخت تصمیم برای ماهی کفال

Figure 6: Predicting the monthly catch of the decision tree for the mullet.

بحث

همانطوریکه در جدول ۱ مشاهده می‌شود، دقت در مدل C&R tree در قسمت آموزشی صد درصد و در قسمت تست ۹۸/۴۲٪ بوده است در مقالاتی که در قسمت پیشینه تحقیق ارائه شده است، به این رقم دقت دست نیافته بودند. در این پژوهش نشان داده شد که این مدل یکی از بهترین روش‌هایی است که برای داده‌های پیش بینی صید استفاده می‌شود. کار جدیدی که در این پژوهش بکار گرفته شده، ادغام کردن داده‌های صید و محیطی برای پیش بینی صید بوده است. حتی در بسیاری از پژوهش‌ها از مدل شبکه عصبی استفاده شده بود که برای داده‌های مورد استفاده در این پژوهش، از این روش نیز استفاده گردید و نشان داده شد که روش درخت تصمیم از شبکه عصبی نیز جواب بهتری گرفته است. مقایسه نتایج و ارزیابی هر دو مدل و همچنین محاسبه خطای مدل‌های شبکه عصبی و درخت تصمیم بر حسب سال صید ماهی کفال در جدول ۴ ارائه شده است.

[Downloaded from isfj.ir on 2024-07-17]

[DOR: 20.1001.1.10261354.1399.29.5.4.0]

جدول۴: مقایسه خروجی مدل شبکه عصبی و درخت تصمیم بر حسب سال صید ماهی کفال

Table 4: Comparison of output of neural network model and decision tree by year of mullet catch

مدل	آموزش	آزمایش	خطای نسبی	میانگین خطای مطلق
شبکه عصبی(NN)	۹۵/۴٪	۹۵/۳۵٪	۰/۱۷۹	۱/۸۱۳
درخت تصمیمC&R tree	٪۱۰۰	۹۸/۴۲٪	۱/۸۷	۱۷/۲۶

در نهایت می‌توان گفت که درخت تصمیم یک روش مناسب برای پیش بینی های درازمدت صید با استفاده از داده های گذشته است و این مسئله به درخت تصمیم خاصی موسوم به درخت تصمیم C&R tree نشان داده شده است. در ضمن، در مسائل پیچیده مانند پیش بینی صید، استفاده از داده‌های صید و داده‌های محیطی در یک منطقه از صیدگاه موجب بهتر شدن نتایج پیش‌بینی می‌شود.

تشکر و قدردانی

با سپاس از کلیه همکاران موسسه تحقیقات علوم شیلاتی کشور که در انجام این تحقیق و نیز آقای دکتر حسن فضلی از همکاران پژوهشکده اکولوژی دریای خزر که در جمع‌آوری داده‌ها ما را یاری نمودند.

منابع

امیری، ک.،۱۳۹۷. مدل سازی عوامل محیطی تاثیر گذار بر میزان صید در واحد تلاش کیلکا ماهیان در سواحل ایرانی دریای خزر با رویکرد سری زمانی. پایان نامه دکتری تخصصی. دانشگاه گیلان. ۳۰۰ صفحه.

بلالی، ب.، ۱۳۹۷. آشنایی با نرم افزار clementine (مفاهیم کلمنتاین) در سایت https://dataacademy.com، ۷۰ صفحه.

خداپرست شیرازی، ج. و صادقی، ز.، ۱۳۹۵. پیش‌بینی تولید آبزیان دریایی در ایران با استفاده از روشARIMA و شبکه عصبی مصنوعی. مجله تحقیقات اقتصاد کشاورزی. جلد ۸. شماره۳.

دریانبرد، غ.، ۱۳۸۸. ارزیابی ذخایر ماهیان استخوانی در سواحل ایرانی دریای خزر (۸۶-۱۳۸۴). موسسه تحقیقات شیلات ایران. ۱۵۸ صفحه.

رضوی صیاد، ب.، ۱۳۶۹. ارزیابی ذخایر و مدیریت ماهیان استخوانی و اقتصادی دریای مازندران .شرکت سهامی‌شیلات ایران. ۱۸۷ صفحه.

عبدالملکی، ش.، غنی نژاد، د.، فضلی، ح. و بندانی، غ.، ۱۳۸۳. ارزیابی ذخایر ماهیان استخوانی دریای خزر در سال ۸۳ـ۱۳۸۲. انتشارات موسسه تحقیقات علوم شیلاتی ایران. ۱۴۵ صفحه.

عبدالملکی، ش.، دریانبرد، غ. و بندانی، غ.، ۱۳۸۴. ارزیابی ذخایر ماهیان استخوانی دریای خزر در سال ۸۴ـ۱۳۸۳. انتشارات موسسه تحقیقات علوم شیلاتی ایران. ۱۴۵ صفحه.

فضلی. ح.، ۱۳۹۷. بررسی ذخایر کفال ماهیان در سواحل ایرانی دریای خزر (۹۶-۱۳۹۴). تهران، موسسه تحقیقات علوم شیلاتی کشور. ۵۲ صفحه.

فضلی، ح. و دریانبرد، غ.، ۱۳۹۶. بررسی تغییرات مکانی صید و گروه بندی شرکتهای تعاونی پره ساحلی در سواحل ایرانی دریای خزر. مجله علمی شیلات ایران. شماره۲. صفحات ۶۸-۶۱. DOI:10.22092/ISFJ.2017.113484

فضلی، ح. و غنی نژاد، د.، ۱۳۸۳. بررسی صید و برخی از جنبه‌های زیست شناختی کفال ماهیان در حوضه جنوبی دریای خزر. مجله علمی شیلات ایران. شماره ۱، صفحات ۱۰۴-۹۷. DOI:10.22092/ISFJ.2004.113719.

نصراله زاده ساروی، ح.، ۱۳۹۷. مطالعه روند سه ساله پارامترهای فیزیکوشیمیایی آب به منظور توسعه آبزی پروری در نواحی مختلف حوزه جنوبی دریای خزر(۸۷-۸۹).

Breiman, L., Friedman, J., Stone, C.J. and Olshen R.A., 1984. Classification and Regression Trees. Taylor and Francis publisher. 368 P.

[Downloaded from isfj.ir on 2024-07-17]

[DOR: 20.1001.1.10261354.1399.29.5.4.0]

Gerami, M.H. and Rabbaniha, M., 2018. Forecasting the anchovy Kilka fishery in the Caspian Sea using a time series approach. *Turkish Journalof Fisheries and Aquatic Sciences*, 18, 1288-1292. DOI: 10.4194/1303-2712-v18_11_05

Ghadirnejad, H., 1996. Population dynamics of grey mullet species (*Liza aurata*) in southern Caspian Sea. Ph.D. Thesis, University of Wales, Swansea. 207 P.

Han, J., Pei J. and Kamber M., 2006. Data Mining: Concepts and Techniques. 800 P.

Hongchun Y., Jintao W., Ying C. and Xinjun C., 2011. Design and implementation of fishery forecasting system based on radial basis function neural network. IEEE Conference on Digital Manufacturing and Automation (ICDMA), pp. 373-376.

Ivanov, V.P., 2000. Biological Resources of the Caspian Sea. KaspNIRKh, Astrakhan. 96 P.

Roger, J. and Lewis, M.D., 2000. An introduction to classification and regression tree (CART) analysis. Introduction to Cart. 14 P.

[Downloaded from isfj.ir on 2024-07-17]

[DOR: 20.1001.1.10261354.1399.29.5.4.0]

Prediction of mullet catch in Iranian waters of the Caspian Sea using decision tree algorithm

Ashournezhad T.[1*]; Shiri M.E.[2]; Rabbaniha M.[1]

[*]tahereh_a2008@yahoo.com

1-Iranian Fisheries Science Research Institute (IFSRI), Agricultural Research Education and Extension Organization (AREEO), Tehran, Iran.
2-Amir Kabir University of Technology, Tehran, Iran

Abstract

Every year the trend of catches and the maximum harvestable water from the water resources are studied and studied. Extensive studies of catch data have been conducted to estimate reserves. The main objective of this study was to predict Caspian mullet catch by decision tree algorithm (C&R tree) presented using Clementine 18 data mining software. Decision tree structure in machine learning is a predictive model that contributes to the observed facts about a phenomenon with inferences about the objective value of that phenomenon. For this purpose, 10-year catch data as a dependent variable (year of catch, month of catch, latitude and longitude of catch point, depth of catch, rate of catch) derived from bony fish stock assessment projects and environmental information (temperature, chlorophyll-a) From the NASA site and the factors of pH, transparency, salinity, oxygen, benthic organisms from northern seas hydrology projects) were considered as independent variables that besides specifying the effective factor on fishing, prediction of fish mullet in water The Caspian Sea also occurs. Among the investigated factors, water surface temperature and salinity were identified as the two main factors affecting fish mullet catchment and the prediction value for 1400, 862/484 tons of mullet with an accuracy of 98/42%.

Keywords: Mullet, Catching, Caspian Sea, Environmental parameters, Decision tree algorithm

[*]Corresponding author

[Downloaded from isfj.ir on 2024-07-17]

[DOR: 20.1001.1.10261354.1399.29.5.4.0]