

Præcision af udbytte forudsigelse ved hjælp af machine learning -teknikker

Denne litteraturgennemgang diskuterer resultaterne af udbytteforudsigelsesteknikker ved brug af machine learning (ML). ML-algoritmer laver forudsigelser ved at finde forbindelser mellem input- og responsvariabler (Crane-Droesch, Andrew, 2018). Artiklerne, der diskuteres i hele, er fra de seneste 5 år, og detaljerede oplysninger om baggrunden, ML-algoritmer, datasæt og resultaterne fra hver artikel kan findes i Excel arket i nedenstående link.

Fremskridt inden for ML har skabt nye muligheder for at forbedre forudsigelsen af høstudbyttet. Resultaterne fra nyere litteratur, der anvender forskellige ML-modeller, viser den vellykkede brug af ML til estimering af udbytte fra byg, hvede og majs. Overensstemmelsen mellem målt og modelleret udbytte er udtrykt kvantitativt ved mean error og mean absolute error (MAE). Resultaterne viste, at ML-modeller kan forudsige bygudbytter med MAE på 6,5 hkg ha⁻¹ og 2,3 – 6,0 hkg ha⁻¹, hvedeudbytter på MAE af 2,9 hkg ha⁻¹ og 70-82% nøjagtighed ved krydsvalidering af modellerne og majsudbytter med MAE af 310,3 kg ha⁻¹ og RMSE mellem 1 and 20%.

Den Gaussian process regression (GPR) estimerede nøjagtigt bygudbytte cirka 1 måned før høst med root mean square error på 7,37 hkg ha⁻¹ and mean absolute på 6,50 hkg ha⁻¹ (Sharifi, Alireza 2021). MAE var lavest fra november til juni og højest fra december til maj, hvilket indikerer højere udbytteforudsigelsesnøjagtighed for modellen med tættere tidsintervaller mellem plantning og høst. Sharifi 2021 fandt ud af, at ud over tidsintervaller, afhænger udbytteforudsigelsesnøjagtigheden også af lokalitet, hvor de højeste bygudbytter i deres forsøg var relateret til fladt land, frugtbar jord og passende vandingsforhold.

Johnson et al 2016 fandt også nøjagtige forudsigelser for bygudbytte ved hjælp af Multiple linear regression (MLR), Model-based recursive partitioning (MOB) og Bayesian neural networks (BNN). Den bedste MLR -model havde en MAE *skill score* på 0,244, den mest præcise BNN-model havde en *skill score* på 0,248, og de bedste MOB-modeller havde en *skill score* på 0,248. Johnson et al 2016 konkluderede, at tilføjelsen af MODIS-EVI (et forbedret vegetationsindeks afledt af Moderat-opløsning Imaging Spectro radiometer) som en forudsigelse, resulterede i en signifikant forbedring af prognosefærdighederne i MLR- og BNN-modellerne, men ikke i MOB-modellerne. Det blev også fremhævet, at den manglende færdighedsforbedring af nonlinear-modeller i forhold til MLR sandsynligvis skyldes det korte (12 år) datasæt, der er tilgængelig for MODIS-data.

ML algoritmer udkonkurrerede lineær regressionsalgoritme med en lavere RMSE i de fleste tilfælde. Random forest (RF) udkonkurrerede andre modeller og gav højere nøjagtighed (RMSE = 0,97) ved forudsigelse af majsudbytte (Khanal et al., 2018). Derudover fandt Petersen, 2020, at selvom Generalized Linear Model (GLM)-modellen fungerede godt, er ML fortsat markant bedre (Generalized Linear Model (GLM) MAE= 3,6 hkg ha⁻¹ og Automated Machine Learning (AML) MAE=2,4 hkg ha⁻¹). I denne Tåstrupanalyse udført af Petersen, 2020 blev det afdækket, at høstudbyttet kan modelleres ud fra integreret RVI, Maksimal RVI, nedbør i juni og reference-evapotranspiration i juni med en afvigelse i MAE på 2,4 hkg ha⁻¹.

Pantazi et al 2016 konkluderede, at SKN-modellen kan bruges til at forudsige hvedeudbytte og klassificere markarealer i forskellige udbyttepotentialezoner. Den gennemsnitlige samlede nøjagtighed ved krydsvalidering for Supervised Kohonen networks (SKNs) var 81,65%, 78,3% for Counter-propagation artificial neural networks (CP-ANNs) og 80,92% for XY-fused networks (XY-Fs) 80,92% (Pantazi et al., 2016). The eXtreme Gradient Boost (XGBoost) model udkonkurrerede andre algoritmer (least absolute shrinkage and selection operator (LASSO) regression, Support Vector Regression (SVR), Random Forest (RF), LSTM network og CNN network) både i nøjagtighed og stabilitet, når man forudsiger majsudbytter (Kang et al., 2020). Kombinationen af "flere estimatorer", "bedre variabler" og "passende tidsfrekvens" resulterede i en forbedring af

nøjagtigheden på op til 5% (XGBoost, 16-day aggregation, med RMSE= 14,8 bushels per acre, MAE= 310,3 kg ha⁻¹ og Mean Absolute Percentage Error (MAPE)= 9%.

Potentialet i mange ML-algoritmer, nemlig RF, GPR, CP-ANNs, MOB, BNN, ALM and SKNs, er signifikant for udbytte forudsigelse af byg-, majs- og hvedeudbytter i forskellige regioner. Resultaterne viste, at ML -modeller kan forudsige bygudbytter med MAE af 6,50 hkg ha⁻¹ og 2,3 – 6,0 hkg ha⁻¹, hvedeudbytter med MAE af 2,9 hkg ha⁻¹ og 70-82% nøjagtighed ved krydsvalidering af modellerne og majs udbytter med MAE af 310,3 kg ha⁻¹ og RMSE mellem 1 and 20%. Tilføjelsen af variabler fra simulation cropping systems model kaldet Agricultural Production Systems sIMulator (APSIM) i alle designede ML-modeller og brug af dem som input til en forudsigelsesopgave kan reducere forudsigelsesfejlmålet med root mean squared error (RMSE) mellem 7 og 20%. Forudsigelser foretaget af hybridmodellen viste mindre bias i forhold til det faktiske udbytte (Shahhosseini et al., 2021). En af de største udfordringer med de fleste ML modeller er, at de ikke i tilstrækkelig grad kunne forudsige udbytter i år med ekstremt vejr, såsom et tørt år.

Referencer

- Crane-Droesch, Andrew. 2018. Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environmental Research Letters*, 13: 114003.
- Johnson, Michael D., Hsieh, William W., Cannon, Alex J., Davidson, Andrew and Bédard, Frédéric. 2016. Crop yield forecasting on the Canadian Prairies by remotely sensed vegetation indices and machine learning methods. *Agricultural and Forest Meteorology*, 218-219: 74-84.
- Kang, Yanghui, Ozdogan, Mutlu, Zhu, Xiaojin, Hain, Christopher and Anderson, Martha. 2020. Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US Midwest. *Environmental Research Letters*, 15, 64005.
- Khanal, Sami, Fulton, John, Klopfenstein, Andrew, Douridas, Nathan and Shearer, Scott. 2018. Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield. *Computers and Electronics in Agriculture*, 153: 213-225.
- Klompenburg, Thomas van, Kassahun, Ayalew and Catal, Cagatay. 2020. Crop yield prediction using machine learning: A systematic literature review. *Computers and Electronics in Agriculture*, 177: 105709.
- Pantazi, X.E., Moshou, D., Alexandridis, T., Whetton, R.L. and Mouazen, A.M. 2016. Wheat yield prediction using machine learning and advanced sensing techniques. *Computers and Electronics in Agriculture*, 121: 57-65.
- Petersen, Carsten. 2020. Foreløbig orientering om forsøg på modellering af høstudbytter i vores langvarige pakkingsforsøg i Taastrup. University of Copenhagen
- Shahhosseini, Mohsen, Hu, Guiping, Huber, Isaiah and Archontoulis Sotirios V. 2021. Coupling machine learning and crop modeling improves crop yield prediction in the US Corn Belt. *Scientific Reports*, 11:1606.
- Sharifi, Alireza. 2021. Yield prediction with machine learning algorithms and satellite images. *Journal of the Science of Food and Agriculture*, 101: 891–896.