

# Predikce Bitcoinu za pomoci rekurentních neuronových sítí

Referát o vědecké publikaci

## 1) Úvod

Referát je realizován formou výtahu z vědecké práce „*Prediction of Bitcoin using Recurrent Neural Network*“, který se zabývá využitím metod strojového učení u problému predikce ceny kryptoměny Bitcoin v porovnání s tradičními metodami pro technickou analýzu finančních produktů.

Bitcoin je decentralizovaná digitální měna, která pro ověření transakcí a uvolnění nových mincí do oběhu nevyužívá centrální autoritu, ale síť zařízení připojených přes Internet a systém Blockchainu.

Bitcoin je stále raný finanční trh. Důsledkem toho je, že tento finanční sektor zažívá velkou volatilitu, která vytváří příležitost k analýze a predikci.

Bitcoin je oproti ostatním kryptoměnám nejstarší a nejpoužívanější a lze snadno získat pomocí těžení, nebo směnou za FIAT měnu na různých obchodních platformách.

## 2) Související práce

### a) Analýza finančních dat

Podle několika strategií se cena na akciových trzích nevyvíjí náhodně, ale je ovlivněna trendy, které doprovází změnu hodnoty. Tyto změny v poptávce mohou být použity pro predikci budoucí ceny.

### b) Analýza časové řady

Analýza časové řady je klasický, poměrně populární postup pro předpověď ceny. Tento přístup funguje velmi dobře, pokud vykazuje řada téměř konstantní trend. V reálném čase, kde ale cena významně fluktuuje není tento předpoklad příliš realistický. Tento přístup ale nemusí spolehlivě fungovat pro predikci na nelineárních datech.

Model random forest je užitečný pro rychlosti výpočtu a schopnost řešit nelinearity v datech pro predikci ceny v reálném čase za využití big data.

Tři modely založené na Bayesovské ANN, lineární regresi a metodě podpůrných vektorů byly zkoumány a bylo zjištěno, že ANN poskytuje z těchto 3 modelů nejpřesnější predikce.

### c) Analýza predikce ceny Bitcoinu

Predikci zítřejší ceny Bitcoinu se několik prací zabývalo za použití několika modelů strojového učení. Strategie využívající soubor několika modelů vyprodukovala zhruba 85 % návrat investice za 50 dnů, čímž překonala strategii trendu předchozího dne, kde byl návrat investice asi 38 %. Poslední strategie založená na predikcích nejlepšího modelu ze souboru vícevrstevných perceptronových sítí zajistila návratnost 53 %.

### d) Závěr ze zkoumaných prací

Metoda random forest a analýza časových řad není pro predikci ceny Bitcoinu pro následující den stabilní, jelikož na Bitcoinovém trhu může nastat výrazný nárůst či pokles. Žádná z analýz ovšem nezkoumala využití konvolučních neuronových sítí a různé kombinace hlubokého učení. Pro zachycení vysokých výkyvů ceny kryptoměn je třeba využít model, který dokáže předvídat manipulace trhu.

### 3) Rekurentní neuronové sítě s dlouhou krátkodobou pamětí

Tři modely byly využity pro předpověď průměrné ceny u 1681 kryptoměn. Dva založené na gradient-boosted decision trees, třetí na rekurentních neuronových sítích (RNN) s dlouhou krátkodobou pamětí (LSTM). Ve všech případech byla sestavena portfolia na základě modelem predikované ceny a u všech modelů byla porovnána návratnost investice. Ze tří modelů dosahoval nejlepších výsledků model založený na LSTM RNN.

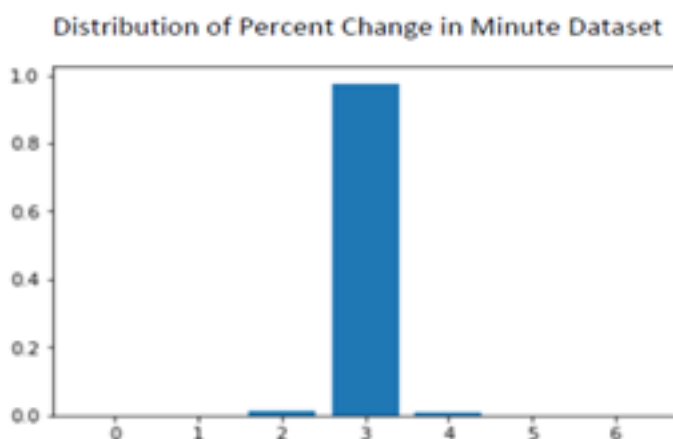
#### a) Datová sada

Datová sada zahrnovala informace o trhu s Bitcoinem (*rozdíl ceny v %, koncová cena, nejvyšší cena, nejnižší cena, vážená hodnota*) získané z portálu *Kaggle two*. Sada dat byla v rozsahu od 1. prosince 2014 do 8. ledna 2018 s granularitou 1 minuta. Každý záznam byl označen štítkem True, pokud hodnota měny vzrostla a jako False, pokud hodnota klesla či zůstala stejná.

#### b) Předzpracování dat

Před trénováním modelu na připravených datech se struktura dat transformuje na 3D tenzor o velikosti okna=50. Poté se absolutní hodnoty dat normalizují.

#### c) Předvídání polarity pomocí RNN



Obrázek 1: Rozdělení procentuální změny hodnoty datové sady

Z obrázku výše vyplývá, že u naprosté většiny záznamů se hodnota měny změnila o 3 % oproti předchozímu záznamu. Cílem bylo vytvořit model, který by dokázal předvídat volatilitu měny v reálném čase. Při poskytnutí dat v řádu minut, nebo sekund model dokázal ještě lépe předvídat inflaci a udržovat si náskok před vývojem trhu.

#### d) Hluboké učení

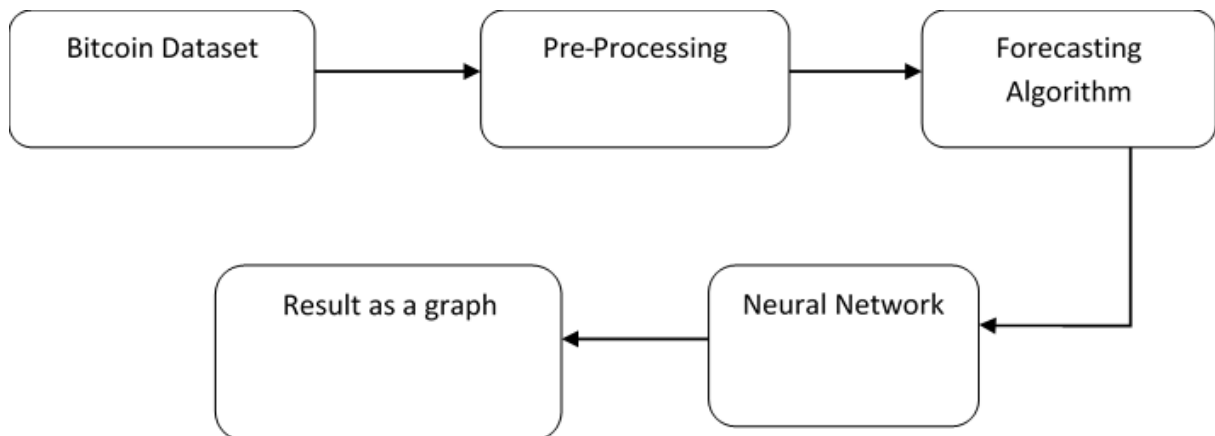
Pro předpověď ceny Bitcoinu na konci předvídaného období, třívrstvá dvousměrná rekurentní neuronová síť byla použita. Jako vstupní data byla použita data ze včerejšího dne.

#### e) Algoritmus

- 1) Data se přečtou z CSV souboru a natrénují se
- 2) Pro  $n$  stromů se zavolá  $n$  mapovacích modulů, které data pro stromy transformují
- 3) Z daného množství dat se vezme 90 % jako trénovací sada
- 4) Data se začnou číst poté co běží mapovací moduly
- 5) Každý mapovací modul zkonstruuje strom a na testovacích datech vyhodnotí predikce
- 6) Testovací data a jejich štítek je předán redukovací jednotce
- 7) Redukovací jednotka spočítá predikované štítky a vrátí ten, který predikuje většina

8) Data jsou zapsána do výstupního souboru

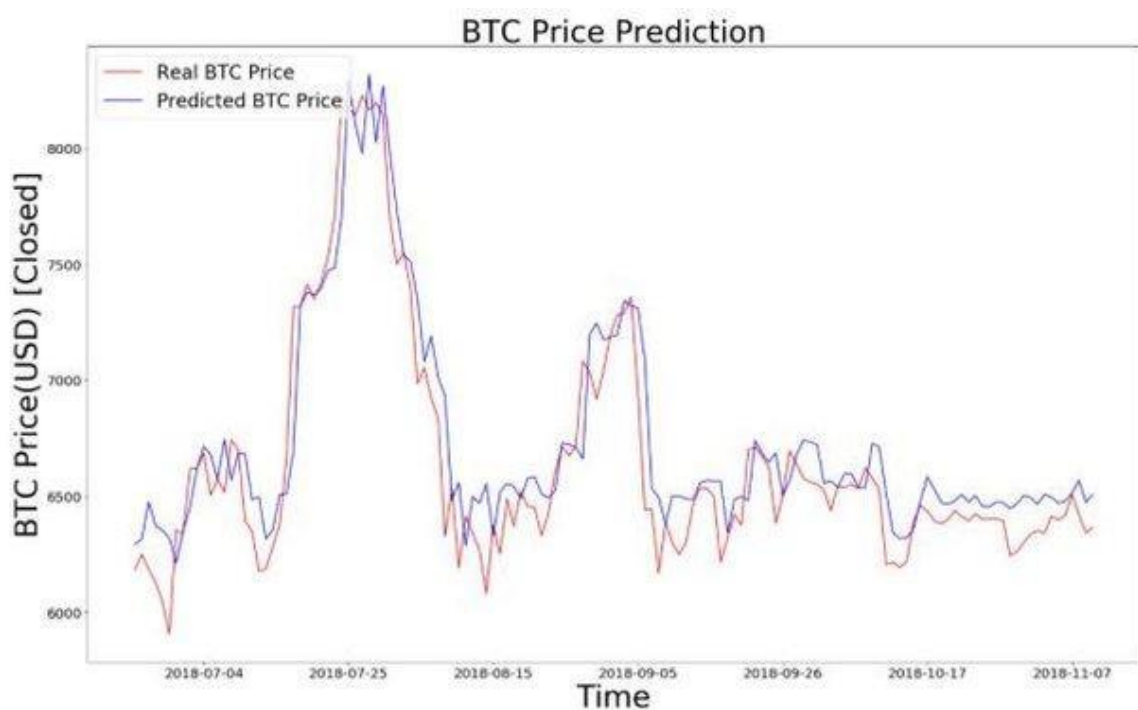
f) Diagram architektury



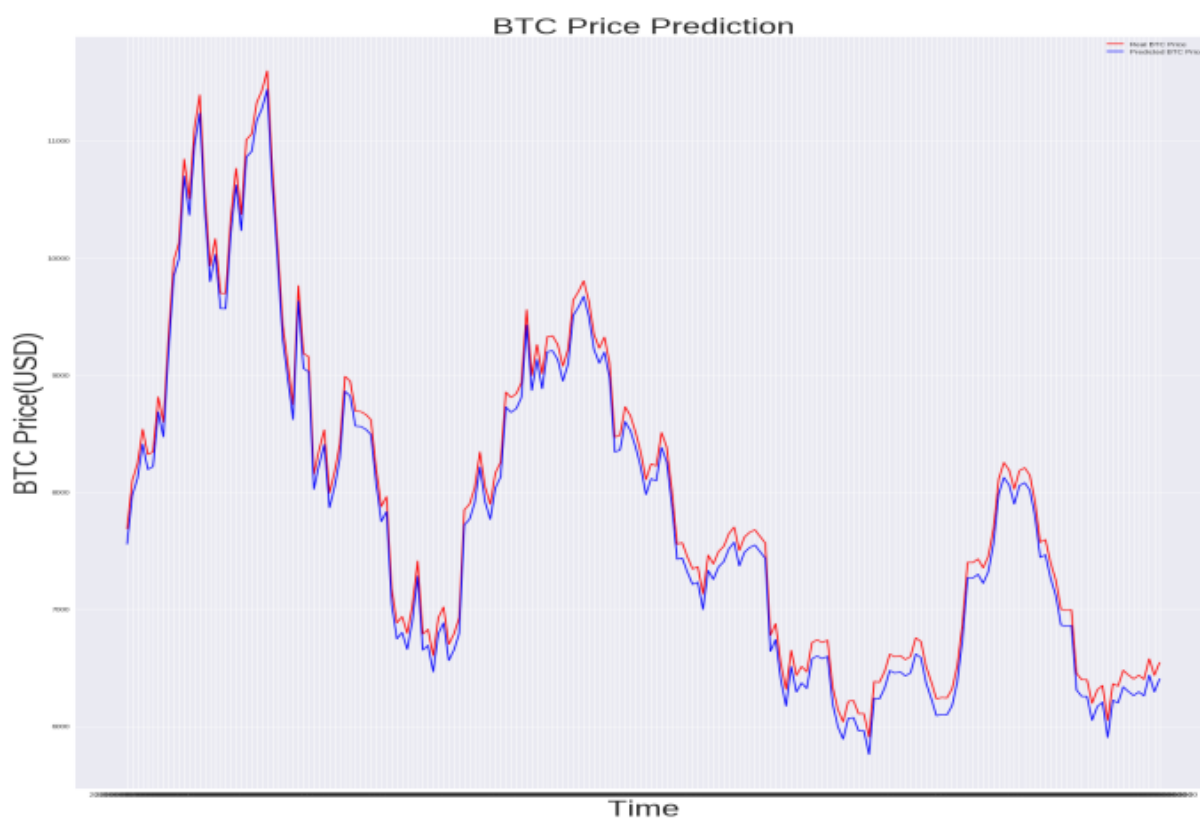
Obrázek 2: Diagram predikce ceny Bitcoinu

Proces predikce kryptoměny Bitcoin probíhá v následujících krocích:

- 1) Načte se vstupní datová sada
- 2) Data se předzpracují
- 3) Použije se predikce štítku pomocí random forest respektive neuronové sítě
- 4) Trénovací a testovací sada dat je porovnávána pro zjištění přesnosti predikce



Obrázek 3: Predikce ceny pomocí random forest



Obrázek 5: Predikce ceny pomocí LSTM RNN

Method	Accuracy
Random Forest	84%
LSTM RNN	92%

Obrázek 4: Přesnost predikce

#### 4) Závěr

Autor nakonec porovnává výkonnost algoritmu random forest a LSTM RNN pro různou velikost datové sady. Jak lze vidět na grafech výše, metoda random forest poskytuje vývoj ceny s přesností 84 %. LSTM RNN poskytuje ovšem přesnost neuvěřitelných 92 % čímž tento model dosahuje nejlepších výsledků ze všech porovnávaných.

Na závěr autor dává příslib dalších studií na tuto tematiku.

#### 5) Zdroje

*Prediction of Bitcoin using Recurrent Neural Network* [online]. SRM Institute of Science and Technology, Kattankulathur, 2020 [cit. 2021-01-10]. Dostupné z: [ijrte.org/wp-content/uploads/papers/v8i6/F7808038620.pdf](http://ijrte.org/wp-content/uploads/papers/v8i6/F7808038620.pdf), ISSN: 2277-3878