

# 1 Základní informace

Jméno projektu	ForeCastle – framework pro distribuovaný výpočet předpovědi prodeje
Zkratka	ForeCastle
Vedoucí	RNDr. Filip Zavoral, Ph.D. <zavoral@ksi.mff.cuni.cz>
Konzultanti	
Anotace	Cílem projektu je vytvořit framework, který bude provádět distribuované výpočty předpovědi prodeje. Framework musí nabízet programátorsky přívětivé rozhraní pro přidávání vlastních matematických modelů předpovědi, možnost importu a správu dat (časových řad popisujících výše prodeje) a v poslední řadě rozhraní pro spuštění výpočtů předpovědi či mechanismus pro hledání nejpřesnější metody předpovědi.

## Motivace

Motivace projektu vychází z reálných požadavků firmy Logio s.r.o., která se zabývá řízením dodavatelských řetězců. To mimo jiné obnáší i předpovědi prodeje. Velký nárůst objemu zákazníků této firmy zapříčinil, že vertikální škálování výkonu už neřeší výkonnostní problémy, se kterými se firma Logio musí potýkat. Cílem tohoto projektu je vytvořit framework, který bude umožňovat distribuovaný výpočet předpovědí (dále forecastů) na více uzlech a umožní tím jednoduchou horizontální škálovatelnost. Framework bude dále přinášet některé nové mechanismy, které aktuální architektura softwaru firmy Logio neřeší. Příkladem je například automatický odhad nejvhodnější metody předpovědi prodeje pro jednotlivé produkty nebo jednoduchá implementace nových předpovědních metod.

## 2 Popis projektu

Forecastem, neboli předpovědí, je myšlen odhad výše prodeje libovolného produktu v budoucnosti v závislosti na výši prodeje v minulosti. Vstupem forecastu je typicky požadovaná délka předpovědi a seznam jednotlivých prodejů, mohou to ale být už tzv. agregovaná data, která popisují celkovou výši prodeje produktu za jednotlivá období (např. den). Pro každou předpověď lze měřit její přesnost, kterou lze získat provedením předpovědi pro již uplynulý časový úsek, pro který máme dostupná data, a poté porovnáním předpovědi se skutečným zaznamenaným prodejem.

## 3 Forecastovací metoda

Forecastovací metodou je myšlen algoritmus, který pro vstupní data vypočítá předpověď. Příkladem takové metody může být například průměrovací metoda, která spočítá celkový (denní) průměr výše prodeje a ten použije jako konstantní předpověď. Dalšími komplexnějšími příklady mohou být jednoduché exponenciální vyhlazování nebo ARIMA model.

## 4 Dekorátory a dekompozitory

Dalšími prvky forecastu jsou dekorátory a dekompozitory. Jedná se o elementy, které upravují vstupní data před vstupem do forecastovací metody. Dekorátorem pak může být myšlen například algoritmus, který upraví data tak, že odhalí a upraví extrémní náhodné prodeje, vlivy sezóny, vlivy denní doby a dne v týdnu nebo jen třeba odstraní náhodný šum. Dekompozitor je pak prvek, který data rozdělí do více složek (trend, výše prodeje, šum...), což poté umožňuje použití dalších technik pro vytvoření předpovědi.

## 5 Spočtení forecastu

Forecastu jednoho produktu se dosáhne aplikací sledu (tzv. forecastovací cesty) jednotlivých dekorátorů či dekompozitorů následované některou forecastovací metodou. Po aplikaci metody je nutné předpověď zpětně opravit o provedené změny (například opět přidat vliv sezóny, který mohl být odstraněn).

Jedním z problémů, který zde vyvstává, je kompatibilita metod a dekorátorů. Ne všechny metody jsou slučitelné se všemi úpravami dekorátorů (například dekorátor může odstranit vliv roční doby, ale metoda ho pro svůj výpočet vyžaduje). Dalším problémem je, že ne všechny prvky tohoto mechanismu umí pracovat se všemi typy dat (seznam prodejů versus agregovaná data). Zásadním problémem ovšem zůstává, že ideální předpověď pro každý produkt může vypadat jinak, čili může sestávat z různých dekorátorů a vybrané metody. Systém proto musí obsahovat mechanismus, který pro daný produkt vytvoří všechny možné validní forecastovací cesty (tzv. forecastovací strom) a změří přesnost každé z těchto předpovědí. Slovem validní je myšleno takové, které sestávají z navzájem kompatibilních prvků. Vítězná nejpřesnější metoda pak bude použita pro budoucí předpovídání prodeje tohoto produktu.

Tento mechanismus je nutné v rámci projektu také vyvinout. Vzhledem k tomu, že měření přesnosti jedné předpovědi je samo o sobě složitá procedura, je vidět, že vypočtení přesnosti celého takového forecastovacího stromu je na výkon velmi náročná aktivita. Zde je jeden z problémů vertikální škálovatelnosti, který je potřeba odstranit.

## **6Rozsah a cíle projektu**

Projekt bude vznikat od úplného začátku, nebude navazovat na žádný konkrétní existující systém firmy Logio. Tato firma žádným takovým podobným systémem nyní nedisponuje. Cílem projektu ovšem není implementace konkrétních forecastovacích prvků (metod, dekorátorů...), ale naopak vytvoření frameworku, který bude podporovat jednoduché přidávání a správu takových prvků. Musí být zpřístupněno kvalitní rozhraní umožňující programátorsky přívětivě napsat nové forecastovací prvky systému. Uživatel-programátor musí mít k dispozici jednoduché nástroje, pomocí nichž definuje požadavky a podobu vstupních a výstupních dat pro tyto nové prvky.

Framework musí zajistit robustní distribuovaný výpočet výše zmíněných úkolů (výpočet předpovědí, stavba stromů, měření přesnosti...). S tím je samozřejmě spojen i problém správy uzlů systému, na kterých výpočet poběží. Jednoduchá horizontální škálovatelnost je jedním z hlavních cílů celého projektu.

Framework musí také podporovat import a správu dat (časových řad prodejů), na kterých výpočet poběží. Počty časových řad produktů se pohybují v desítkách až stovkách miliónů a systém na to musí být připravený i z hlediska úložiště. Importovaná data budou přicházet v různých formátech, které budou blíže specifikovány. Systém musí podporovat jednoduché přidávání importovacích modulů, které budou zdroje dat obsluhovat.

Framework bude obsahovat webové uživatelské rozhraní, pomocí kterého bude možné sledovat stav systému (běžící procesy, seznam výpočetních uzlů...), spravovat data a forecastovací prvky, zobrazit výsledky výpočtů a spustit vlastní výpočty.

## **7Některé další požadavky**

Dalším požadavek se týká specifikace uzlů systému, na kterých poběží výpočty. Cílový operační systém bude Linux. Na hardwarovou stránku nejsou kladeny žádné speciální požadavky, půjde o běžné serverové stroje či počítače. Konfigurace přidávaných uzlů by měla být minimální, ne ovšem automatizovaná. Systém by měl dále rozlišovat uživatele a jejich oprávnění k přístupu k jednotlivým prvkům systému.

## Platforma, technologie

- Jedním z požadavků firmy Logio je cílový jazyk, ve kterém musí být systém vyvíjen. Tím je jazyk C#.
- Cílová platforma uzlů, na kterých výpočty poběží, je operační systém Linux a běhové prostředí Mono.
- Jako middleware pro meziprocesovou komunikaci bude pravděpodobně zvolen RabbitMQ.
- Relační databáze pro uložení metadat byla zvolena MariaDB.
- Jako úložna samotných dat časových řad bude kvůli jejich velkému objemu a požadované škálovatelnosti NoSQL databáze MongoDB.
- Back-endová část webového rozhraní není zadavatelem specifikována. Zvolili jsme ovšem C# a ASP.NET (MVC a Web API).
- Front-endovou část bude tvořit webová aplikace postavená na frameworku AngularJS.

Výše uvedené technologie se mohou změnit v závislosti na analýze technologií, která bude jako součást projektu provedena.

## 8 Odhad náročnosti

Na projektu se bude podílet 5-6 studentů. Práce na projektu bude po schválení probíhat dle následujícího hrubého harmonogramu s předpokládaným termínem dokončení na jaře 2016:

### 1. Analýza a specifikace požadavků

Sběr a analýza požadavků klienta, průzkum dostupných forecastovacích technik

### 2. Analýza technologií

Průzkum možností realizace distribuovaného systému, analýza existujících technologií pro distribuované počítání, analýza dostupných databázových systémů a úložišť dat

### 3. Návrh architektury systému

Návrh informační architektury, distribuovaného řešení, jednotlivých procesů uvnitř systému, datového modelu, programátorského a uživatelského rozhraní systému

### 4. Implementace “Proof of Concept”

Implementace distribuované části řešení, která demonstruje použití zvolených technologií

### 5. Implementace zbylé části systému

Implementace forecastovacích mechanismů, programátorského a uživatelského rozhraní systému. Během implementace budou průběžně vytvářeny i automatické testy

### 6. Vytvoření dokumentace

Bude vytvořena uživatelská i programátorská dokumentace

## Vymezení projektu

*Projekt je zaměřen na následující oblasti (zaškrtněte vyhovující):*

Diskrétní modely a algoritmy	
	diskrétní matematika a algoritmy
	geometrie a matematické struktury v informatice
	optimalizace
Teoretická informatika	
	Teoretická informatika
Softwarové a datové inženýrství	
X	softwarové inženýrství
X	vývoj software

X	webové inženýrství
X	databázové systémy
X	analýza a zpracování rozsáhlých dat
Softwarové systémy	
	systémové programování
	spolehlivé systémy
X	výkonné systémy
Matematická lingvistika	
	počítačová a formální lingvistika
	statistické metody a strojové učení v počítačové lingvistice
Umělá inteligence	
	inteligentní agenti
	strojové učení
	robotika
Počítačová grafika a vývoj počítačových her	
	počítačová grafika
	vývoj počítačových her

## 9Poznámky

*žádné*